The Dissertation Committee for Melissa Christine LoPalo
certifies that this is the approved version of the following dissertation:

Essays on the Determinants of Worker Productivity and Labor Market Outcomes

Committee:

Sandra B Youngblood, Co-Supervisor

Dean E Spears, Co-Supervisor

Michael L Geruso

Diane L Coffey

Essays on the Determinants of Worker Productivity and Labor Market Outcomes

by

Melissa Christine LoPalo

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2019

Acknowledgments

I am indebted to Sandy Black, Mike Geruso, and Dean Spears for their support, encouragement, and advice, and for being constant sources of inspiration. I also wish to thank Marika Cabral, Diane Coffey, Leigh Linden, and Tom Vogl for generously sharing their knowledge, time, and expertise. I am grateful to David Beheshti and Kevin Kuruc for reading countless drafts and sharing ideas. Finally, I wish to particularly thank my friends and family, who provided invaluable support during my graduate education.

Essays on the Determinants of Worker Productivity and Labor Market Outcomes

Melissa Christine LoPalo, Ph.D.
The University of Texas at Austin, 2019

Supervisors: Sandra B Youngblood
Dean E Spears

This dissertation examines determinants of worker productivity, labor market outcomes, and population health.

The first chapter, previously published in the Journal of Public Economics, examines the impacts of cash assistance on refugee labor market outcomes. I exploit variation across states and over time in the generosity of cash assistance available to refugees upon arrival in the U.S. and study the impacts on wages and employment. I argue that cash assistance is randomly assigned to refugees conditional on characteristics such as education and country of origin, as refugee placement is decided by a committee that does not meet with the refugees or learn their preferences. I find that refugees resettled with more generous cash assistance go on to earn higher wages, with no significant change in the probability of employment. The effects are largest for highly-educated refugees.

The second chapter examines the impact of temperature on the productivity and job performance of outdoor workers in developing countries. I overcome data challenges with studying individual-level productivity by studying household survey interviewers as workers. Using data from Demographic and Health Survey interviewers in 46 countries, I find that

interviewers complete fewer interviews per hour worked on hot and humid days, driven by an increase in working hours. I also find evidence that suggests that workers allocate their effort towards tasks that are more easily observed by supervisors on hot days.

The third chapter, previously published in Social Justice Research and co-authored with Diane Coffey and Dean Spears, examines the role of social inequality in population health outcomes in India, focusing on the case of casteism and child height in India. We describe evidence from the India Human Development Survey showing that children in villages with more strongly casteist attitudes are shorter on average, an association that is statistically explained by the association between casteism and the prevalence of open defecation.

Table of Contents

List of Tables

List of Figures

Chapter 1

The Effects of Cash Assistance on Refugee Outcomes[1]

1.1 Introduction

The United States accepted approximately 70,000 refugees in 2015, incurring initial resettlement costs of $1.1 billion with the goal of helping refugees achieve self-sufficiency as quickly as possible. In addition to resettlement assistance for basic needs such as rent and food upon arrival, refugees have access to several forms of cash assistance during the resettlement process. Refugees who come with families are eligible for Temporary Assistance for Needy Families (TANF) for up to five years, while refugees who meet the eligibility requirements for TANF but do not have access to the program for categorical reasons, such as not having any children, are eligible for up to eight months of Refugee Cash Assistance (RCA), which is designed to match TANF benefit levels in most states.

However, poverty rates among refugees remain high relative to natives. According to a recent report, 44 percent of refugees lived in households with family incomes below 200 percent of the poverty line in 2009-2011, compared with 33 percent of natives (Randy Capps and Kathleen Newland, 2015). The same report found that reliance on public assistance was higher among refugees than both natives and economic immigrants. In the general debate over public assistance programs such as TANF, policymakers generally weigh the efficiency

[1] This paper was published in the Journal of Public Economics, as Melissa LoPalo (2019).

consequences of disincentivizing labor against the consumption smoothing benefits of this temporary social insurance. Indeed, there is evidence that access to welfare programs such as TANF among non-refugees generally results in lower probability of employment and lower wages for the affected populations (see, for example, Robert A. Moffitt (2003)).

Refugees, however, are a unique population. Refugees often arrive in the U.S. with few resources at their disposal, but many arrive with significant human capital. Refugees are as likely as U.S. natives to have a college degree, but median household incomes are lower than both natives and other immigrants (Capps and Newland, 2015). Given their unique educational and income profile, refugees may interact differently with welfare programs than natives and other immigrants, potentially shifting the typical efficiency-equity tradeoff.

Refugees in the U.S. offer a unique opportunity to examine the impact of cash assistance on liquidity-constrained, college-educated individuals entering a new labor market. Average treatment effects of cash assistance across the whole welfare receiving population may mask substantial heterogeneity for individuals who do not fit the average profile of a welfare recipient: newly-arrived, highly-educated immigrants may have a particularly large amount to gain from cash assistance. In addition, despite the differences between refugees and average welfare recipients, refugees present a unique opportunity to examine the long-term impacts of a temporary welfare program; the vast majority of them use some form of cash assistance in the first years after arrival, making it possible to examine the long-term impacts of cash assistance despite the lack of data on eligibility and welfare use upon arrival.

Finally, there is little evidence on the effect of the refugee resettlement program on labor market outcomes. Performance measures within the refugee resettlement program are structured to reflect the priority of short-term self sufficiency: the Office of Refugee

Resettlement (ORR) requires resettlement agencies to track employment retention and wages for refugees, but only for 90 days after resettlement, making it difficult to track the effect of the overall resettlement program on longer-term outcomes. In this paper, I provide the first econometric analysis evaluating the effect of access to cash assistance in the years following resettlement on longer-term refugee labor market outcomes. In particular, I examine the effects of wage generosity rules in likely refugees' state of residence at the time of their arrival on wages and other outcomes in household survey data.

I use variation across states and over time in TANF generosity, measured simply as the maximum benefit level for a family of three in a given state, to evaluate the effect of government assistance on the outcomes of likely refugees, since most refugees receive cash benefits based on TANF rates upon arrival in the U.S. Notably, this means that the levels of benefits most refugees receive were set for the population as a whole, since TANF is a population-wide program and not specific to refugees. Therefore, variation in TANF levels is arguably exogenous to voters' and policymakers' attitudes towards refugees.

Using decennial Census and American Community Survey data and the sample of refugees resettled since the inception of the TANF program in 1996,[2] I estimate substantial positive wage effects of 5-8% per $100 of cash benefit levels among employed refugees. The strongest effects accrue to the most highly educated refugees, whose wages increase by approximately 8-11%. I find that the types of jobs that refugees work in change in response to TANF: the mean educational level of refugees' occupations increases, suggesting that refugees get better jobs following more generous cash assistance. In contrast, I do not

[2]I exclude refugees settled under the Aid to Families with Dependent Children (AFDC) program, which preceded TANF, due to the lack of data on benefit levels by state for several years in the period 1980-1996.

find significant results on employment or use of food stamps. I test for heterogeneity in wage effects on several additional dimensions, finding further evidence that the wage effects accrue to refugees who are more primed for labor market success: the effects are larger for refugees with higher English ability and fewer children, for example. In addition, event study difference-in-differences plots around the timing of changes to state TANF policies show that labor outcomes for refugees follow a parallel path prior to increases in generosity and then diverge for refugees settled in the periods after the increase, supporting the underlying assumptions of the empirical strategy.

Overall, the results provide suggestive evidence that, while the welfare system largely treats refugees the same as other low-income populations, they may have a higher marginal benefit from additional cash assistance than other recipients. In fact, the size of the coefficients suggest that the positive impacts on wages rival the impacts of interventions targeted specifically to improve labor market outcomes of low-income populations, such as training and workforce development programs (Michael Greenstone and Adam Looney, 2011). The magnitude of these results suggest that there may be substantial potential for positive effects of cash assistance among underemployed populations in the U.S. In addition, by focusing on short-term labor market outcomes of refugees, analysis of the resettlement program may be overlooking an important effect of cash assistance on long-term wages. Furthermore, the findings lend empirical support to the idea—pointed out in past Government Accountability Office (GAO) reviews—that the emphasis on short-term outcomes can come at the cost of the provision of services that may improve refugee labor market outcomes in the long run, such as helping refugees that arrive in the U.S. with foreign credentials to overcome barriers to using their skills in the U.S. (GAO, 2012).

1.2 Motivation and Previous Literature

This paper contributes to the prior literature in several ways. The literature on TANF's predecessor, AFDC, generally finds that higher benefit levels reduce labor supply and increase caseloads (Moffitt, 2003). In addition, the sizable literature on welfare reform in the U.S., which introduced time limits, work requirements, and greater state ownership of welfare programs, among other changes, generally finds that the reforms were associated with decreased caseloads, increased employment, and increased wages (Rebecca Blank, 2002). Similar results have been found among immigrants as well (Robert Kaestner and Neeraj Kaushal, 2005). However, refugees are quite different from the overall population of welfare recipients who are the focus of these studies. Most of the evidence on the effects of welfare participation on labor market outcomes in the U.S. necessarily focuses on recipients with low educational attainment: in fiscal year 2015, only 7.5 percent of adult TANF recipients had more than a high school education (HHS, 2016). Given that 28 percent of refugees have a bachelor's degree or higher, refugees present a unique opportunity to examine the impact of welfare participation among beneficiaries with high levels of human capital.

In addition, very little evidence exists on the impacts of cash assistance on the longer-term outcomes of adults who use it. Studies that have examined longer-term effects have also tended to find negative impacts, with David J. Price and Jae Song (2018) finding that adults who used cash assistance worked in less prestigious jobs in the long term, while Elizabeth T. Wilde, Zohn Rosen, Kenneth Coutch and Peter A. Muennig (2014) find some evidence of increased mortality hazards.

There are several mechanisms through which cash assistance could affect labor market outcomes such as wages and employment in the long run. It is well understood that many highly-educated immigrants are underemployed in the United States, partially because of obstacles to practicing their professions such as re-certification requirements (Capps and Newland, 2015). For example, immigrants who were nurses in their country of origin must pass a standardized examination and a professional evaluation of their credentials and prove their English proficiency, along with potentially passing additional nursing courses before practicing in the U.S. (ORR, 2012). This means that refugees wishing to use advanced degrees in the U.S. often have to invest in U.S.-specific human capital such as English proficiency and American certifications upon arrival (Kalena E. Cortes, 2004). Cash assistance may help recipients overcome liquidity constraints to partake in labor market outcome-enhancing investments.[3]

Additional aid could also provide a cushion that allows refugees to hold out for a better job match in their initial search, leading to improved labor market outcomes. Alternatively, and especially given the income requirements to remain on TANF and RCA, welfare receipt may crowd out work. Each of these effects implies longer periods of decreased employment among newly arrived refugees, but the welfare implications are different.[4] Finally, refugee resettlement agencies generally provide aid with the employment search process, and fast employment is a priority. These services could also affect labor market outcomes in the long run without causing delayed employment, although, as I discuss in Section 3, funding for these services is generally allocated based on refugee arrivals and is therefore not likely to be correlated across states and over time with cash assistance.

[3]In Appendix A.0.1, I use a simple framework similar to (Cortes, 2004) to derive the prediction that U.S.-specific human capital investments may increase if refugees receive higher welfare benefits upon arrival in the U.S.

[4]See Raj Chetty (2008) for a theoretical framework examining a similar tradeoff in the context of unemployment insurance.

I provide evidence on the welfare impacts of the resettlement program by examining the effect of cash assistance generosity on the long-run employment status of refugees and the wages of employed refugees.[5] As publicly available data on outcomes such as recertification do not exist, these outcomes provide the best possible evidence of the effect of the resettlement program on the ultimate success of refugees in the labor market. Furthermore, I am able to observe heterogeneity in the effect by educational attainment, allowing me to examine whether the effect is concentrated among refugees who likely had qualifications to practice highly-skilled professions in their home countries.

This analysis also contributes to our general understanding of the determinants of refugee success in the U.S. While this is the first study to examine the relationship between welfare assistance for refugees and their labor market outcomes, previous studies have used similar identification strategies to examine other determinants of refugee success, in particular network effects within refugee communities. The existing literature on refugees tends to make the argument that placement of refugees is exogenous and looks at network effects within refugee communities. Per-Anders Edin, Peter Fredriksson and Olof Aslund (2003) argue that refugees in Sweden are effectively placed randomly conditional on observable characteristics, since they are not able to choose where they live within Sweden. They focus on a period when the housing market in Sweden was particularly tight, which meant that refugee placement was largely based on the availability of housing in different areas of Sweden. They find that living in enclaves is associated with an earnings gain, particularly for high-income ethnic groups. Olof Aslund, Per-Anders Edin, Peter Fredriksson and Hans Grönqvist (2011)

[5] Evidence from the Annual Survey of Refugees suggests that by 4 years after arrival, the employment rate of refugees matches that of the whole U.S. population (ORR, 2016).

use a similar identification strategy to investigate the effect of living in a neighborhood with adults of the same ethnicity on the school performance of immigrant children in Sweden. They find that a standard deviation increase in the share of highly-educated members of one's own ethnicity improves one's GPA by 0.9 percentile ranks.

A similar argument that placement of refugees is random conditional on observables can be made in the U.S., since refugees do not meet with the committees that decide their placement nor do they express their location preferences. Lori Beaman (2012) uses data from a voluntary agency that resettles refugees in the U.S. to examine the effects of social networks on labor market outcomes, making the assumption that placement is random controlling for refugee characteristics. She finds that outcomes improve with the number of tenured refugees in one's social network, while they deteriorate with the number of refugees settled in the same year or prior year, perhaps due to competition. I also make the assumption that placement of refugees is random conditional on observable characteristics, but I focus on the effect of government cash assistance on the labor market outcomes of refugees resettled in the United States, a topic that to my knowledge has not yet been covered in the literature. This question has direct implications for the optimal design of a refugee resettlement program.

Unlike Beaman (2012) I do not observe the refugee status of the individuals in my dataset. I follow several papers in the U.S. context in using household survey data to identify likely refugees. George J. Borjas (2000) uses 1980 and 1990 decennial Census data to compare the effects of living in ethnic enclaves for refugees, identified by country of origin, and economic migrants. Cortes (2004) also uses 1980 and 1990 Census data and examines the differences in characteristics and outcomes of refugees and economic immigrants that arrived in 1975-1980. She finds that refugees initially had lower earnings and worked fewer hours but by 1990 earned more, worked more, and spoke better English compared to economic immigrants. Like Borjas (2000) and Cortes (2004), I identify likely refugees based on country of origin, but I make additional restrictions to my sample based on year of arrival.

1.3 The "Lottery Effect": Refugee Resettlement in the U.S.

1.3.1 Placement Decisions

After receiving refugee status, newly arrived refugees are matched with one of nine voluntary resettlement agencies that provide services to refugees.[6] Importantly, the committee does not meet the refugees before deciding their placements, and refugees have no channel for expressing location preferences. Refugees with relatives already living in the U.S. are likely to be placed near them.

For refugees without ties in the U.S., the resettlement agencies consider factors such as language capacity in the host community, average rent, public assistance rates, employment opportunities, and existing ethnic communities in making a match (Refugee Council, 2015). Given that the committee has some biographical information on refugees and is aware of state budget trends and economic opportunities in potential settlement sites, it is vital that my identification strategy control for any refugee characteristics that may influence placement. I explore this issue in detail below. Resettlement agencies also consider the number of refugees already resettled in a community, often placing refugees in areas that have had success with refugee resettlement in the past (GAO, 2012). The refugee is then transported

[6] The refugees are allocated to the resettlement agencies in "round-robin fashion according to... approved percentages" (Refugee Council, 2015) at weekly meetings of representatives of each agency. Each organization has offices throughout the U.S., and the largest states have sites run by most, if not all resettlement agencies.

to the community chosen for them out of about 190 that currently resettle refugees.

In the past, refugees came from a relatively small number of countries, but the group has become increasingly diverse in recent years, with the implication that fewer are placed near family and more in a relatively random location (Andorra Bruno, 2011). Placement of refugees who have family in the U.S. is arguably less random, but this fact will not bias my estimates unless the location of family members is systematically related to welfare generosity in different states. This would occur only if either placement of the first wave of refugees from a certain sending country was not random conditional on observables, or if endogenous secondary migration occurred among the first arrivals from that country. Both concerns are addressed to the extent possible in the identification strategy outlined below.

1.3.2 Resettlement Program

A majority of refugees enroll in either TANF or RCA upon arrival: in 2015, 40.3 percent of refugee households that had been in the U.S. for less than one year reported using TANF benefits, while 24.9 percent used RCA (ORR, 2016). Use of these programs declines sharply with years in the U.S.; only 6.4 percent of households that had been in the U.S. for five years reported still using TANF. Benefit levels for RCA are generally based on the TANF formula for that state. Levels in states with public private partnership or Wilson/Fish programs,[7] however, vary from the amount suggested by the TANF formula to a higher amount approved by the ORR (GAO, 2011). Given that my empirical specification assumes that all refugees are exposed to the variation in TANF benefits across states, the effect of

[7] These states are Oregon, Alaska, Idaho, Nevada, Colorado, Texas, Oklahoma, North Dakota, South Dakota, Minnesota, Louisiana, Alabama, Kentucky, Maryland, Massachusetts, and Vermont.

the inclusion of states with higher-than-predicted benefit levels should be to attenuate the estimated effects on labor market outcomes. I present results excluding these states in a robustness check.

TANF rates, in turn, vary significantly by state, creating part of the phenomenon known as the "lottery effect" in refugee placement (GAO, 2011). As of July 2016, maximum benefit levels for a single-parent family of three varied from $170 per month in Mississippi to $923 per month in Alaska, with a median benefit level of $431 (Ife Floyd and Liz Schott, 2015). The variation persists even after accounting for differences in costs of living. Figure 1.1 shows the maximum monthly benefit for a family of three in 2015 in each state after adjusting the dollar amounts using the Bureau of Economic Analysis (BEA) Implicit Regional Price Deflator (IRPD).

After controlling for differences in costs of living by state, the maximum benefit levels vary from about $180 to $626 per month, providing significant cross-sectional variation for use in identification. Cost of living data by state is not readily available for my entire sample period (the BEA index runs from 2008 to 2015), so I am unable to control for any changes in cost of living in my regression framework.[8] However, I control for state fixed effects, which will absorb any time-invariant differences in cost of living, and year of immigration fixed effects, which will control for any common trends in costs of living across states. Specifications with state-specific trends will account for any linear changes within

[8]In Table A6 I test the robustness of my results to using the 2008-2015 IRPD data to adjust my measure of TANF generosity. In this test, I fit a linear trend in the IRPD data for each state and extrapolate the trend backwards to 1996 and forwards to 2016. The results are not statistically significantly different, though the magnitude on the effect on wages in my preferred specification is slightly larger. This is to be expected given some of the cross-state variation in benefit levels I use for identification is not attributable to actual differences in generosity.

the states.

Furthermore, there has been significant variation in both absolute and relative generosity of the program by state over time since its inception as a replacement for the Aid to Families with Dependent Children (AFDC) program in 1996. For example, in July 1996, Wyoming was the 27th most generous state, offering $360 per month, while in July 2016 it had risen to 5th, offering $652. The real value of TANF benefits has fallen in many states since 1996. The differential extent to which the generosity declined was due in several cases to differing policies by state for adjusting the benefits over time rather than to actual changes in welfare policy. For example, Texas adjusts its benefit levels each year in proportion to changes in the federal poverty level, while Wyoming adjusts according to changes in the state's cost of living index. Meanwhile, 15 states have not adjusted the nominal level of their benefits since 1996. There have also been instances of states adjusting benefit levels due to changes in the business cycle or voter preferences, among other reasons. For example, several states, including the large refugee resettlement states of California and Washington, cut cash assistance benefits during the Great Recession (Liz Schott and Ladonna Pavetti, 2011). Since business cycle conditions also likely affect the outcomes studied, it is important to control for them. I therefore include controls for the unemployment rate in the resettlement state both in the year the refugee arrived and in the year of interview.

Figure 1.2 shows changes in the real value of TANF maximum benefit levels for a family of three in the nine largest states for refugee resettlement in my sample.

It is clear that a good deal of the variation over time occurs due to the erosion in the real value of benefits resulting from states opting to not adjust the nominal benefit levels. However, there are multiple instances of sharp rule changes, as well. Specifications using only state, year of interview, and year of immigration fixed effects will use the erosion in real values of benefits as a source of variation over time, while specifications using state time trends will use it to a lesser degree, since any linear changes in determinants of the outcome variables over time will be parsed out. Figure 1.3 shows variation over time in TANF benefits for the same states, but with state-specific linear time trends removed. Specifically, the figure shows the residuals from regressions of TANF benefits on a variable for year. The figure shows that the majority of the variation left over after removing state time trends comes from a few states, especially California, New York, and Illinois.

In fiscal year 2015, over 3.1 million households used TANF benefits, suggesting that refugees constitute a very small portion of TANF beneficiaries and are unlikely to influence the policymaking process. The significant variation both across states and over time in welfare benefit levels for a program not explicitly designed to serve refugees presents an opportunity for examining the effects of welfare generosity on refugee labor market outcomes without concern about endogeneity of benefit levels with respect to refugee characteristics. Although outcomes for refugees may vary with many other features of the communities in which they are placed, such as the quality of public transportation, public attitudes towards immigrants and refugees, and services offered by local branches of resettlement organizations, examining the effects of variation in welfare generosity gets at one crucial aspect of the "lottery effect."

States have many degrees of freedom in designing TANF laws beyond setting benefit levels (Moffitt, 2003). These include differential time limits, for which I control in a robustness check, as well as other aspects such as work exemptions, asset limits, and other eligibility requirements. While refugees are especially likely to be eligible for TANF due to their low income and lack of assets upon arrival in the U.S., they are likely to be exposed to variation in welfare rules beyond what is represented using my simple measure of generosity.[9] Furthermore, a minority of refugees use other welfare programs such as Supplemental Security Income (SSI) and General Assistance in applicable states which also vary in generosity by state. In my preferred specification, however, these other policies would have to be changing in a non-linear manner at the same time as benefit levels to bias my estimates.

In addition, although most refugees who receive cash assistance from the ORR obtain it through RCA, a portion receives money through the Matching Grant program, which is a federal program that provides matching funds to resettlement agencies to offer cash assistance for four to six months. In fiscal year 2009, 31 percent of refugees who received cash assistance from the ORR (this does not include refugees using TANF or SSI) used the Matching Grant program (GAO, 2011).[10] The effect of the Matching Grant program should be to attenuate the estimates of the effects of TANF generosity, since some of the refugees in my sample will essentially not be treated, and also because the association between number of slots and benefit levels will effectively dampen the variation across states in generosity.

[9]For example, refugees are eligible for food stamps as well, and food stamp benefits are calculated based on income, which includes TANF income. This results in food stamps dampening the variation in assistance generosity. In results not presented here but available upon request, I use total SNAP and TANF maximum benefits as the treatment variable. The results are very similar.

[10]The exact level of the cash benefits is decided by the resettlement agency but must exceed $200 per month. An ORR employee indicated in an interview that the agencies do not vary the level of Matching Grant benefits by state of placement, but that they do tend to allocate their Matching Grant "slots" to areas with the highest demand. TANF benefit generosity could be one factor determining demand for Matching Grant slots. While I do not have data on number of Matching Grant slots by state for every year in my sample and am thus unable to directly control for any differences in slot numbers, I do have data for several years from the ORR. In results not presented here, I compare number of slots to TANF benefit levels for 2010-2014, finding a correlation of -0.016. This indicates that a generous TANF program does predict a lower number of Matching Grant slots, but the association is weak.

Refugees also have access to services through resettlement agencies upon arrival to the U.S. These organizations provide case management services, such as aid with initial housing, clothing, and transportation needs and with enrolling any children in nearby schools. They also provide employment services and help refugees find English classes. Resettlement agencies receive funding to provide these services through the ORR according to the number of refugee arrivals. In addition, the ORR allocates money through grants to states, resettlement agencies, and local resettling communities through several additional programs, such as the Targeted Assistance Program (TAG), which provides additional funding to states with large refugee populations, high unemployment, high secondary migration, or high rates of public assistance use. This funding is not correlated in the cross section across states with TANF maximum benefit levels. Therefore, while these types of services may have significant impacts on refugee labor market outcomes, they are unlikely to represent a major pathway for the results I estimate in this paper.

1.4 Data and Summary Statistics

My main dataset comes from the 2000 decennial Census Public Use Microdata files (PUMS), which provide a 5 percent random sample of the United States, as well as the 2001-2016 American Community Survey (ACS) data, which replaced the long-form Census. The first few years had smaller sample sizes, but since 2005 each year of ACS data has provided a one percent sample of the U.S. My sample is composed of probable refugees (details in section 4.1) aged 25 and older, who are likely to have completed their formal education by the time I observe them. I exclude refugees who arrived before the TANF program began in 1996.

Summary statistics on major control variables are presented in Table 1.1 for different sending populations in the 2012-2016 ACS data. Mobility rates vary significantly by sending region, with some of the more recent waves, such as Southeast Asia, Africa, and the Middle East, having higher rates of mobility. This may suggest that refugees reshuffle somewhat in the first few years after arriving before settling down more permanently in a state. The ORR Annual Reports to Congress generally mention that secondary migration of refugees mainly occurs in the first year after resettlement (ORR, 2016). Indicators of economic success vary substantially by sending population as well, and it appears that older sending populations are on average more successful in terms of wages, again consistent with either cohort effects or assimilation.

Summary statistics on the same variables are presented in Table A1 for the full sample of refugees (above the age of 24) as well as economic migrants, defined as the foreign born who do not fall into my refugee sample, in the 2000 Census and 2012-2016 ACS, respectively. According to this dataset, about 2.1 percent of refugees have moved between states in the last year. In the 2016 ACS data, about 2.4 percent of the population as a whole had moved between states in the prior year. This suggests that refugees are not a particularly mobile group once they arrive in the U.S., which is reassuring evidence in favor of the assumption that endogenous secondary migration isn't a major problem for the identification strategy. Table A1 also shows evidence of possible cohort effects and/or assimilation among the refugee population; refugees observed in 2012-2016 have higher wages, employment rates, and labor force participation rates and are more likely to speak English than refugees observed in 2000. This could occur if more recently arrived refugees have different skill distributions from older sending populations, or it could reflect improved outcomes among populations with greater tenure in the U.S. The latter would occur if the sample of refugees observed in 2012-2016 had greater experience in the U.S. due to a decreased arrival rate. Of course, the changes could also occur due to changes in labor market or other conditions for refugees between 2000 and 2012-2016. In addition, the 2000 Census sampled a larger proportion of the U.S. population than is reflected in the 2012-2016 numbers, suggesting that the sample means in 2000 may be closer to the population average.

1.4.1 Identification of Refugees

As mentioned previously, I identify likely refugees by country of origin and year of arrival, as neither the Census nor the ACS ask the foreign born whether they are refugees or economic immigrants. Previous papers such as Cortes (2004) and Borjas (2000) identify refugees based on country of origin; their selections are not time-varying. I introduce a methodology for selecting probable refugees based on fluctuations in arrival numbers over time; that is, I classify a country as a refugee country only in years in which numbers of refugee arrivals are high relative to the number of total arrivals (including economic immigrants). I use data from the Yearbook of Immigration Statistics, which is put together by the Department of Homeland Security, on refugee arrivals and immigrant visas issued by region and country of nationality from 1996 to present to identify refugee sending countries.[11] If refugees make up more than 60 percent of total arrivals, defined as refugee arrivals plus immigrant visas issued, for a certain country-year, immigrants arriving in that year that

[11] The sending countries in my sample are: members of the former Soviet Union, Bosnia and Herzegovina, Croatia, Bhutan, Burma (Myanmar), Laos, Afghanistan, Iraq, Cuba, Congo, Eritrea, Liberia, Somalia, and Sudan. See the Appendix for more details on sample construction.

were born in that country are identified as likely refugees.[12] For example, arrivals from the Balkans during the 1990s are coded as likely refugees, while immigrants from Mexico are not. My methodology is similar to a Migration Policy Institute (MPI) report that identifies immigrants from country-years when refugee admissions exceeded 40 percent of the estimated foreign born population identified in the ACS as probable refugees (Capps and Newland, 2015). Given that my identification strategy relies on the precise timing of the arrival of the refugee, I focus on refugees per immigrant in a given year of arrival instead. Finally, to limit variation in the sample from small changes in actual arrival numbers, I use a five-year centered moving average of the fraction of refugees.

Figure 1.4 shows the percent of refugees in the sample of refugees arriving since 1996 residing in each state. About 50 percent of the refugees in the sample reside in three states: California, Florida, and New York. This compares to around a quarter of the population 25 and over as a whole, according to the 2016 ACS. Furthermore, about 72 percent of refugees live in the top 10 receiving states. This compares with about 54 percent of the general population. The amount of geographic concentration among refugees is similar to that of the non-refugee foreign born population in my sample, though the state rankings differ slightly; refugees are much more concentrated in Florida, while relatively smaller populations reside in California and Texas, for example. The concentration of refugees is perhaps not surprising given that historical refugee populations tend to be reinforced by the arrival of family reunification cases and the fact that placement decisions are made taking into consideration local resources.

[12] I examine the robustness of my results to higher thresholds below.

Figure 1.5 shows the distribution of refugees across states compiled from arrivals data by state from the ORR from 1996-2015. Summed across years, this data gives the distribution that would have occurred if all refugees stayed in their original state of resettlement. Comparing Figure 1.4 and Figure 1.5 jointly provides a check on my method for identifying refugees and on the extent of secondary migration of refugees in the years of my sample. The patterns are broadly similar; however, there are a few differences in the distribution worth noting. The ORR data suggest that about 21 percent of refugees were resettled in Florida between 1996 and 2015, while about 36 percent in my sample live in Florida. In contrast, California, Texas, and Minnesota, among others, have relatively fewer refugees in my sample than in the ORR data.

Secondary migration data is available for select years during the sample period from the ORR (2012-2015;1995-1998). The secondary migration data back up the increased concentration of refugees in Florida, but they also suggest that Minnesota has received very large inflows of secondary migration on average, for example. The differences in the distribution I observe in the data from the one implied by the ORR data on initial resettlement and net secondary migration may in part be a consequence of the relative difficulty in picking up refugees from smaller sending countries using my method. For instance, Minnesota has large populations of refugees from various African countries, such as Somalia. Given that the total sending populations are smaller from these areas than from countries such as Cuba or the Soviet Union, my measure of how many live in a certain state may be relatively noisy.

A related but different issue is that my method for identifying refugees will likely classify some economic migrants as refugees. For example, there are many Vietnamese living in the U.S. who arrived as refugees, but there are also many more who arrived through other channels, such as through family-based immigration (MPI, 2015). In Table A2, I show the average (smoothed) fraction of refugees issued for each country in my sample.[13] The proportions vary widely by sending country, with Croatia having an average refugee share of about 0.65 and Bhutan having a ratio of 0.94.[14] If this misclassification occurs randomly with respect to my identifying variation in welfare generosity, then it may be expected to bias my estimates downwards, since some of the individuals that I identify as refugees will not be exposed to the "treatment" of interest.

1.4.2 Welfare Generosity Measure

For data on TANF benefit levels and other features of the TANF programs, I use the Welfare Rules Database. The database was put together by the Urban Institute and allows users to make queries on various aspects of TANF programs, such as activities exemptions, dollar amounts, time limits, and income eligibility tests. For my baseline measure of welfare generosity, I use maximum monthly benefits for a family of three,[15] filling any gaps in information with Center on Budget and Policy Priorities data (Floyd and Schott, 2015). The Welfare Rules Database also contains information on time limits; most states allow beneficiaries to be on TANF for up to five years, but a handful of states, including Connecticut,

[13] The denominator is refugees plus immigrant visas.

[14] Note that the immigrant totals are not comprehensive: I give only immigrant visa numbers, not including migrants on nonimmigrant (temporary) visas.

[15] Among refugees aged 25 and older in my sample who have at least one child, the median family size is over four for both the whole sample and for the recently arrived. However, the median family size is three for the refugee sample overall. In the main results, I focus on benefit level for a family of three because it is a standard metric for reporting welfare generosity, allowing me to fill in any gaps in information in the Welfare Rules Database and cross-reference using other sources. However, I test the robustness of my results to using the rules for a family of four in Table A7. The results are qualitatively unchanged, though the magnitudes are smaller.

Arizona, Arkansas, Florida, Georgia, and Michigan, have adopted shorter time limits. The shortest is Connecticut, at 21 months. Given that the eligibility time limits may be important in determining the responsiveness of refugee outcomes to welfare money, I exclude any variation in welfare generosity coming from changes in time limits as a robustness check.

1.5 Empirical Methodology

I exploit variation both across states and over time in TANF benefit levels to estimate the effect of being randomly (conditional on observable characteristics) settled in a state with a certain level of generosity on outcomes, accounting for time-invariant state characteristics as well as year of interview and arrival fixed effects. Specifically, I estimate the following equation on the 2000-2016 repeated cross section data:

$$y_{isTt} = \beta G_{sT} + \gamma X_{it} + \delta UR_{sT} + \alpha UR_{st} + \theta_s + \xi_t + \eta_T + \epsilon_{isTt} \tag{1.1}$$

where y_{isTt} is the outcome of refugee i in year t who lives in state s and arrived in the U.S. in year T and X_{it} is a vector of characteristics of the individual that may influence placement outcomes. I control for country of origin,[16] number of children, marital status, educational attainment in four categories,[17] and English ability,[18] as well as age and its square.

The need to control for biographical data that may have an effect on placement decisions presents a challenge for identification of the effect of welfare generosity on labor market

[16]The countries are listed in Table A2.

[17]Less than high school, high school, some college, and college and above.

[18]I create a dummy variable for English ability, coding those who respond that they either don't speak English or don't speak English well in the Census or ACS form as not speaking English, and those who respond that they speak English well or very well or only speak English as speaking English.

outcomes if some of these characteristics themselves respond to welfare generosity. For example, if educational attainment influences allocation of refugees to more or less generous states, but refugees have a tendency to increase their educational attainment after arrival in response to welfare generosity, then it will be difficult to disentangle the effect of welfare on labor market outcomes from the effect of the increased educational attainment on labor market outcomes. I explore this issue more carefully below.

I also control for business cycle conditions in both the year of arrival and the year of interview using the unemployment rate (UR).[19] To the extent that the resettlement agency committee responds to local business cycle conditions in making placement decisions or state legislatures respond to business cycle conditions in setting TANF benefit levels, these time-varying state characteristics may not be exogenous to the placement decision or welfare generosity. θ_s, ξ_t, and η_T are state, year of interview, and year of arrival fixed effects, respectively, allowing me to parse out any time-invariant characteristics of a particular state that may affect refugee outcomes, such as general attitudes towards refugees. I am unable to simultaneously control for year of interview and year of arrival fixed effects and the number of years the refugee has been in the U.S. In Table A10 I test the robustness of my results to removing year of interview fixed effects and adding controls for number of years in the U.S. The results are virtually unchanged. The coefficient of interest, β, measures the effect of welfare generosity in terms of TANF maximum benefit levels for a family of three in the refugee's state of residence at the time of his or her arrival, represented by G_{Ts}. I cluster standard errors at the state level.

[19]I downloaded unemployment rate data from the Federal Reserve Economic Data (FRED) database, which uses Bureau of Labor Statistics data on unemployment.

I estimate Equation 1.1 using several outcome variables available in the data, but my main results look at the effects of welfare generosity on log earnings; I estimate the effect on yearly earnings excluding refugees who do not have positive wages.[20] Using yearly earnings allows for the possibility that some of the effects occur through refugees adjusting labor supply along the intensive margin.

1.5.1 Support for Identifying Assumptions

As mentioned previously, the committee that makes placement decisions has the pertinent information on cash assistance levels and trends by state, so it is critical to my identification strategy that the committee not be sending refugees with certain unobserved characteristics that predict their success to more generous states, that the levels of states' cash assistance do not themselves respond to refugee placement decisions, and that other state-level correlates of refugee outcomes are not changing at the same time as cash assistance levels. I examine these assumptions more carefully in this section, and Section 7 contains further discussion of alternative explanations.

1.5.1.1 Refugee Characteristics and Placement Decisions

One of the identifying assumptions is that placement is exogenous conditional on observable characteristics, or in other words, that refugees are not systematically sorted into more or less generous states based on unobservable characteristics. There is no method for directly testing whether this identification assumption is valid (Esther Duflo, Rachel Glen-

[20] All annual earnings are deflated to real 2016 dollars using the Bureau of Labor Statistics Consumer Price Index for Urban Wage Earners and Clerical Workers (CPI-W).

nerster and Michael Kremer, 2008). However, the specifications minimize the potential bias from endogeneity of placement by controlling for state fixed effects. If the refugee resettlement committee does sort more needy refugees into more generous states, by, for example, taking into account a broader range of lottery effect characteristics such as public transportation and public attitudes, then as long as the perception of these state characteristics do not change with TANF benefit levels over time, the state fixed effects should parse this out. Furthermore, in my preferred specification, with the addition of state linear time trends, confounding factors would have to change nonlinearly at the same time as TANF benefit levels to bias my estimates.

However, it is clear from documentation of the placement process that resettlement agencies do take into consideration state trends in cash assistance levels as well as other current local economic conditions in making placement decisions. Therefore, it is crucial that I am able to adequately control for any information about the refugees that resettlement organizations have when making their placement decisions. I am unable to observe some information on refugees that the agencies have when making their decision, but fortunately, they do not meet the refugees that they are placing before the decision is made, so the number of unobservable characteristics that could influence placement outcomes is limited. Resettlement organizations do have access to some medical data and information on religion and ethnicity that I do not have, though region of origin controls may capture a good part of the variation from the latter two characteristics. It is also somewhat reassuring that if anything, assuming that the agencies direct needier cases to communities with better resources, endogeneity of placement should bias me against finding any positive wage results from increased welfare generosity.

One way to indirectly examine the assumption of exogenous placement conditional on observable characteristics is to test the extent to which resettlement agencies appear to use characteristics observed by the econometrician to make placement decisions. To this end, in Table 1.2 I show the results of Equation 1.1 run without controls for refugee characteristics, using refugees' observable characteristics as the outcome variables. The regressions all include state, year of immigration, and year of interview fixed effects, as well as linear state time trends. This table provides a rough test of the extent to which placement officers use refugee characteristics in making placement decisions.[21] The majority of characteristics show no significant relationship with cash assistance generosity, and those that do are sensitive to the choice of specification. Using the baseline sample, it appears that refugees with less than a high school education and those from Southeast Asia are more likely to receive generous cash assistance. In column 2 of Table 1.2, I widen the sample of refugees to those who came from a country in a year when at least 40 percent of immigrants were refugees, to rule out the possibility that the lack of correlations in the table are driven by a lack of statistical power. In this specification, refugees with more children get significantly more assistance, as do those from the Middle East. It is reassuring that in general, observable characteristics do not appear to be strongly correlated with the level of cash assistance. In addition, to the extent that they are correlated, it appears that it is the less skilled populations that receive more money.[22]

[21] It is worth noting that these characteristics are observed at the time of interview, not in the year of immigration. Therefore, some of them may have changed between those two points in time, making them a potential outcome variable.

[22] The results of this table fit well with conclusions derived from interviews with International Rescue Committee (IRC) officials involved in placements that are described in Beaman (2012). These interviews implied that officials used few biographical details of refugees to make placement decisions.

1.5.1.2 Benefit Levels and Other Determinants of Outcomes

As with any differences-in-differences empirical design, this empirical strategy relies on a "parallel trends" assumption. In this case, the paper relies on the assumption that if no state changed their TANF benefit levels during the sample period, refugee wages would have evolved similarly in all states. Lacking data on refugee wages in this counterfactual scenario, we can't directly test this assumption. We can test, however, whether trends in refugee wages before major TANF changes tend to match the states that did not experience the change. Figure 1.6 shows the results of one such test for parallel pre-trends. This figure displays the results of a version of Equation 1.1, where the independent variables of interest are indicator variables for whether the refugee arrived in a state 2, 3, or more than 4 years before or 0, 1, 2, 3, or 4 years after an increase in TANF maximum benefit levels of at least 3 percent, allowing us to compare wages for refugees who arrived before vs. after a major TANF hike. The comparison group in this regression is refugees who arrived one year before a large increase. Refugees in states that never experienced a large TANF increase are excluded.[23]

The coefficients on wages for refugees arriving before a TANF increase are statistically insignificant, providing evidence consistent with the parallel trends assumption. However, it is important to note that refugees arriving in the years immediately before a TANF increase are not uniformly unaffected by the increase, since refugees can be on TANF for up to five years. Several factors may cause the lack of effect for refugees arriving in the years leading up to a large increase. One is that cash assistance use falls precipitously in the year after arrival,

[23]In Figure A3 I show the results of a specification where these refugees are included in the omitted category with refugees who arrived one year before a large increase. The results look very similar.

with TANF use falling by half and RCA use virtually disappearing, according to the 2015 Annual Survey of Refugeess. Further, cash assistance may be most important in determining outcomes in the year after arrival, when refugees are first settling into the new labor market, although this is difficult to investigate empirically. Finally, refugees arriving the year before a TANF increase are arriving in the year of a small TANF decrease on average, which could work as a headwind against finding a positive effect of the TANF increase in the year after arrival. In contrast, as expected, the coefficients for the year of an increase and thereafter are positive, though not always statistically significant. This provides suggestive evidence that TANF changes, and not other determinants of outcomes trending differentially across states, are the cause of the main effects.

To delve further into the parallel trends assumption, I run similar tests for non-refugee outcomes to look for correlations between TANF changes and overall labor market conditions. In Figure A1, I examine how native wages in the year of interview evolve around large TANF increases. That is, I compare natives interviewed before and after a TANF increase with the excluded group of natives interviewed one year before the increase. There is some indication that native wages are higher a few years before the increase and lower a few years after, indicating that TANF increases may be related to the business cycle. However, the coefficients after the increase are negative, suggesting that if anything, the positive refugee effects are biased downwards by a negative correlation between TANF and overall wages.[24] I run similar tests for non-refugee immigrant wages in Figure A2. The coefficients for years before a TANF change are again small and statistically insignificant. However, a negative association appears to develop several years after the change.

[24]In results not presented here but available upon request, the employment pattern looks similar.

1.6 Results

Results from Equation 1.1 are shown in Table 1.3. The effect per \$100 of TANF monthly benefit for a family of three is exhibited in the first row. The specifications all include state, year of arrival, and year of interview fixed effects as well as state time trends. The first two columns do not include controls for demographic characteristics or local economic conditions; the controls are added in columns 3 and 4. Columns 2 and 4 exclude refugees who have been in the U.S. for fewer than five years, as these refugees may still be using TANF benefits. The coefficient on TANF generosity for log annual wages is relatively stable across specifications; these results suggest that refugees experience around 5.2-7.7% higher wages per \$100 of maximum monthly benefit level. As expected, there are substantial wage premiums for educational attainment and English ability (the coefficients on the educational categories are relative to the omitted category of college and above).

Results from Equation 1.1 run separately for men and women are presented in Table 1.4. These specifications include state, year of interview, and year of immigration fixed effects and state time trends. The results suggest that men's wages may be more sensitive to welfare generosity than women's, but the difference is not statistically significant. Thus, it is difficult to draw definitive conclusions about any heterogeneity in response to welfare generosity by gender.

The next two tables explore some potential channels through which the effect of welfare generosity on labor market outcomes might operate. Table 1.5 presents the results from specifications allowing for each category of educational attainment to have a different interaction effect with the measure of welfare generosity. Since the main and interaction effects for college-educated refugees are omitted, the coefficient on welfare generosity can be interpreted as the effect for college-educated workers. These results strongly suggest that the effects are largest for highly-educated refugees. This is intuitive, as highly-educated refugees may reasonably have the most to gain from the additional liquidity up front if it allows them to wait for a better job match or invest in U.S.-specific human capital such as English proficiency, which is strongly correlated with higher wages in all specifications, or recertification.

As mentioned previously, it is possible that educational attainment could also respond to welfare generosity. To minimize this "bad control" problem, in columns 3-4 I present results from the same specifications, but excluding refugees who arrived before the age of 25. It is possible that refugees who arrived after this age respond to welfare generosity by increasing educational attainment, but it is plausibly less likely. The results are little changed, suggesting that educational attainment may not be an important outcome variable. In Table A5, I examine heterogeneity in the main effect by several other refugee characteristics. I find that cash assistance is most helpful to refugees with better English ability, refugees with fewer children, and older refugees. This reinforces the idea that cash assistance may be particularly helpful for refugees that are already most primed for success in the labor market. It's worth noting, however, that the effect of cash assistance on wages is positive and statistically significant for nearly every group, including those with less than a high school education and those who do not speak English.

As I mentioned in Section 2, welfare generosity may affect wages through increased investment in U.S.-specific human capital or through an improved job match, either of which would result in the refugee ultimately finding a higher-paying job. One way to test whether one of these mechanisms is occurring is to look at the effect of welfare generosity on refugees' occupational characteristics, which I examine in Table 1.6. I investigate the impact of the TANF generosity measure on occupational mean wage and occupational mean education in years. A positive effect on occupational mean wage or occupational mean education would suggest that refugees placed with higher cash assistance went on to work in better jobs. In columns 1 and 2 the outcome variable is mean education in years for all workers (both native and foreign born) 25 and over in the refugee's occupation in 1990 and 2000, respectively. In columns 3 and 4, I average the hourly wage of all individuals working in each of the occupations identified in the Census and ACS and assign that wage by occupation to each refugee. In each of these specifications (columns 1 through 4), the coefficient on the TANF benefit level is positive, and the effects are statistically significant for occupational mean education as well as occupational mean wage using the 1990 calculation, providing significant evidence that refugees are, in fact, getting better jobs when they are resettled with higher levels of cash assistance.

Table A4 further explores the question of whether refugees invest more in their human capital upon arrival if they have additional liquidity from welfare benefits. Here, I examine the effect of removing controls for English ability (column 1) and education (column 2). If the coefficient on TANF generosity increased after removing these controls, it might suggest that some of the effects of generosity on wages were occurring through the channel of improved English proficiency or increased educational attainment (or that these characteristics were important determinants of placement outcomes). Removing controls for education increases the magnitude of the coefficient, but the difference is not statistically significant, consistent with the conclusions of Table 1.5. Similarly, the magnitude of the main result does not change with the removal of the control for English ability.[25] Of course, these results do not rule out the possibility that refugees increase their U.S.-specific human capital in ways not testable using this data: they could be completing re-certification requirements or attending training on practicing their professions in the U.S. context, for example. More research is needed to definitively draw conclusions on the mechanisms for the positive effects on wages that I find here.

In Table 1.7, I examine the effects of welfare generosity on a few alternative outcome variables: probability of employment (columns 1 and 5), use of food stamps (columns 2 and 6), probability of moving between states in the past year (columns 3 and 7), and probability of speaking English (columns 4 and 8). The fact that I find no evidence of an effect of welfare generosity on employment on the extensive margin may be the result of one of any number of factors particular to refugees. For one, newly arrived refugees may have fewer resources, making it difficult to live solely on benefits from welfare programs and pushing them to work if they are able. In addition, resettlement agencies often tie the provision of services (such as employment services) to compliance with employment and training rules, including an obligation to accept any job offers unless there is an acceptable reason to decline. Finally, cash assistance is temporary unless the refugee qualifies for SSI, so refugees I observe with significant tenure in the U.S. likely work out of necessity. However, my results only apply to employment effects along the extensive margin; it may be the case that refugees adjust their hours based on available cash assistance. In addition, it's possible that refugees do prolong their job search in response to more generous cash assistance, but that this effect does not

[25] In the next table, I run a specification with the English dummy as the outcome variable, and do not find significant effects of welfare generosity on probability of speaking English.

last long enough to be detected using these data. Columns 2 and 6 show no significant evidence of an effect on the use of food stamps, though the point estimate in column 6 is negative. For secondary migration, though the effect of cash assistance in the full sample is statistically significant at the 10 percent level, most of the coefficients on welfare generosity are not statistically significant, and most of them are close to zero. This provides some reassuring evidence in support of the identifying assumption that endogenous secondary migration is limited, though it only gives evidence on the prior year. Finally, columns 4 and 8 provide additional evidence that English ability is not an important outcome variable. If English ability responded to the level of cash assistance, it would provide evidence that increased English language skills could be a major mechanism for the increased wages we observe.

In order to further understand the finding of no effects on employment, I test for heterogeneity in employment effects across a number of observable characteristics. In Table 1.8, I show the results of specifications that include interaction effects between TANF maximum benefit level and a number of refugee characteristics. No group of refugees in these specifications shows a statistically significant effect of cash assistance on probability of employment. However, in some cases, the effects on employment are statistically significantly different for different groups of refugees. For example, while neither English-speaking nor non-English-speaking refugees have an effect of cash assistance on employment, the effect on English speaking refugees is statistically significantly more positive.[26] The same is true

[26] In results not presented here but available upon request, I test for interaction effects along several other characteristics, including age, number of own children, and marital status. The interaction effect with age is statistically significant, but very close to zero. None of the rest of the interaction effects are statistically significant.

for educational attainment: the effect is statistically significantly more positive for more highly-educated people. This reinforces the idea that cash assistance is especially helpful to highly-educated refugees in the labor market. However, the interaction effects are uniformly quite small, never larger than about 2 percent. In Section 7.1, I examine the potential for selection into employment in more detail.

1.7 Robustness Checks

In this section, I present several robustness checks to examine the consistency of the results using alternative subsamples and specifications. First, as mentioned before, some states have programs that allow more flexibility in terms of whether they prefer to set RCA rates according to their TANF formulas. The first column of Table 1.9 presents estimates from Equation 1.1, but excluding these states. These specifications include state, year of interview, and year of arrival fixed effects as well as state time trends. The coefficient on log wages is virtually the same as in earlier specifications. If the coefficient on log wage were larger in this specification, it would have been consistent with the idea that refugees in the excluded states are less affected by the identifying variation in welfare generosity.

As a second robustness check, I estimate Equation 1.1 excluding refugees from Cuba. Given that this population is exceptionally geographically concentrated (84 percent in my sample live in Florida), it is reassuring that the results appear to be robust to dropping it, though the main coefficient marginally loses statistical significance. Finally, in column 3, I exclude any variation in welfare generosity coming from changes in time limits for TANF. As mentioned previously, most states use the standard time limit of five years for TANF. While a handful of states have shorter time limits, only a few experience changes in these time limits during my sample period, so the vast majority of variation in generosity coming from time limits is absorbed by state fixed effects. However, Arizona, Delaware, Kansas, New Mexico, South Dakota, and Montana experience changes in time limit policies according to the Welfare Rules Database during my sample period. Less than three percent of my sample of refugees live in these states, but to exclude the possibility that these changes in generosity are driving my results, I remove these states from my sample in these specifications.[27] The results are virtually unchanged.

In Figure 1.7 I examine the robustness of my results to experimenting with different cutoffs for identifying likely refugees. The estimate furthest to the left in the figure presents the results of Equation 1.1 run on the sample of all foreign born individuals, including refugees. Moving to the right, I narrow down the sample so that the cutoff criteria for a refugee sending country becomes more stringent. The second estimate shows results with immigrants from country-years where the fraction of refugees exceeded 10 percent identified as refugees. Moving to the right, each estimate increases the cutoff fraction by 10 percentage points. The size of the point estimates generally increases as the share of refugees in the sample grows, until the fraction reaches 0.7. The point estimates then become more volatile as the standard errors increase, but there appears to be a downward trend. Referring to Table A2, this is perhaps not surprising once we note that sending countries with lower refugee shares tend to have more highly educated refugees and the results are stronger among more highly educated individuals. The share of the sample that is college educated is plotted on

[27]In results not presented here, I attempt to control flexibly for these time limits, allowing for an interaction effect between maximum benefit level and time limit. However, given the lack of variation in time limits picked up in my estimation strategy, the results are not well identified. Therefore, while time limits and other TANF design details may represent important dimensions for the effect of welfare generosity on welfare outcomes, my identification strategy is not well suited to examining them.

the second y axis.[28]

I also present evidence that a similar effect of TANF generosity does not exist for the non-refugee foreign born.[29] The Personal Responsibility and Work Opportunity Reconciliation Act of 1996 mandated that non-refugee immigrants who have been in the U.S. for fewer than five years are not eligible for federal funding under welfare programs such as TANF and SSI, though many states have since implemented programs to extend TANF benefits to newly arrived immigrants using state funding. Furthermore, none of the non-refugee foreign born population is eligible for RCA.

Table 1.10 provides evidence that the positive effects I find on wages apply uniquely to refugees. The effect of welfare generosity is negative, either indicating a causal negative effect of TANF generosity on wages or an effect of an omitted variable that is correlated with the non-random location decisions of immigrants. As I show in Table A9, immigrant characteristics, in contrast to refugee characteristics, are significantly correlated with cash assistance benefit levels in the main specification, indicating that estimates of effects for immigrants are unlikely to be causal.[30]

In Table A11, I show the results of triple difference specifications using a variant of

[28]In Table A8, I present the main results using a lower cutoff of 40 percent. The magnitudes are smaller, as expected given that the effect is intent to treat, but they remain statistically significant in the preferred specifications.

[29]Controls for the sending regions in this specification are: Mexico, Canada, Latin America excluding Mexico, Northern and Western Europe, Eastern Europe, East Asia, Southeast Asia, Southwest Asia and the Middle East, Africa, and Oceania, rather than the refugee sending countries.

[30]In Table A12, I look at heterogeneity in both the wage and employment impacts of TANF generosity on non-refugee immigrants, finding evidence that the negative impacts are concentrated among non-English speaking and low-skilled immigrants. This reinforces the idea that interactions with the welfare system are different for high human capital individuals, although again, the results among non-refugee immigrants are not well identified.

Equation 1.1 where non-refugee immigrants are used as a control. In each specification displayed in the table, refugee status is interacted fully with all controls, including individual controls, fixed effects, and state-specific time trends. In all specifications, the interaction between refugee status and TANF maximum benefit level is positive and statistically significant, while the main effect on TANF maximum benefit level is negative, indicating that the positive effect of cash assistance generosity is specific to refugees. Columns 3 and 4 show the coefficient on interactions between refugee status and educational attainment and English ability, showing that the returns to these types of characteristics vary strongly between refugees and non-refugees.

Finally, in Table A13 I show the results of Equation 1.1 run on a sample of native workers with a high school education or less. I examine the impact of cash assistance levels in the year of interview on their wages in the same year as a check on whether high cash assistance levels are associated with good outcomes for all workers, rather than the effect being particular to refugees. Columns 1 and 2 show that while the impact of cash assistance is estimated to be positive, it is small and statistically insignificant. Meanwhile, columns 3 and 4 show that the impact on employment is close to zero.[31]

As a final check on the sample definition for refugees, I estimate Equation 1.1 using a few additional sample definitions; the results are shown in Table A14. Column 1 shows the results for the baseline sample definition, which uses a five-year centered moving average of the fraction of refugees entering in a year from a certain country to identify likely refugees.

[31] These results are in no way meant to represent a causal estimate of the effect of TANF programs on native workers' labor market outcomes. I use a simplified welfare generosity measure better suited to refugees than to native workers, and there is of course no reason to think that location of native workers is exogenous.

Column 2 shows the results when the fraction is not smoothed; they are virtually unchanged.

The fact that most of the variation over time in cash benefits comes from just a few states once state time trends are removed raises the concern that the results may be driven by policy changes other than changes in cash assistance levels in those particular states. In Figure 1.8, I estimate Equation 1.1, dropping each of the 15 top states for refugees in my sample one by one. In all of these specifications, the coefficient remains nearly identical. Given that the wage results are robust to the removal of all of the states where the refugee population is mostly concentrated (nearly 81 percent of the sample lives in these states), we can gain confidence that they are not driven by an unrelated state-specific policy change that affected refugee outcomes.

1.7.1 Alternative Explanations

1.7.1.1 Selection into Employment

This paper focuses on wages conditional on employment as the primary outcome of interest. While these specifications produce economically and statistically significant results, the possibility remains that the effects are driven by selection rather than a true causal effect. I previously showed that there was no statistically significant effect of cash assistance on employment, but if the average treatment effect masks significant effect heterogeneity on observable or unobservable characteristics, it remains possible that positive selection into the sample of employed refugees could be contributing to the results. If certain high-ability refugees are more likely to be employed if they are resettled with more cash assistance, for example, we may see average wages among the employed go up even if individual refugees do not earn higher wages than they would have in the counterfactual scenario where they were resettled with less cash assistance.

As I noted earlier, there is some evidence of heterogeneity along observable characteristics, though the effects of cash assistance on employment are not statistically significant or sizable among any individual group. However, the presence of statistically significant interaction effects warrants closer attention to this alternative explanation for the results.

While it's not possible to directly decompose the effect on wages conditional on employment into a selection effect and a pure wage effect, in Table A15 I perform one indirect test for the presence of a selection effect. The table contains the results of a version of Equation 1.1 run without controls, using refugee characteristics as the outcome variables of interest. This table is similar to Table 1.2, except that here the specifications are run only on the sample of employed refugees. The results of the table provide some evidence on how much cash assistance generosity changes the composition of employed refugees. Like in Table 1.2, very few refugee characteristics produce statistically significant results. Column 1 contains the full sample of employed refugees while column 2 excludes those arriving in the most recent five years. Depending on the specification, there's some suggestion that more generous cash assistance draws more low-educated refugees, fewer English speaking refugees, more female refugees, or refugees with more children into employment. These groups generally have lower than average wages, suggesting that it's unlikely that selection accounts for positive conditional wage effects.[32]

[32] As a final test for selection effects, I estimate a Heckman two step specification (James J. Heckman, 1976), which takes into account that the econometrician observes wages of refugees only when they are employed, and specifies that employment is a function of observable characteristics of refugees and some error term, which may be correlated with the error term in the wage equation. The results of this specification are in Table A16, which suggests that there is a small yet significant degree of selection (ρ and λ are statistically significant), but that this cannot account for the effects of cash assistance on wages.

1.7.1.2 Cash Assistance Generosity and Refugee Populations

It is possible, in theory, that state legislatures may respond to the placement of certain types of refugees in their state by adjusting cash assistance levels. Given that refugee cash assistance is tied to the generosity level of TANF in the majority of states, and refugees make up a tiny portion of overall TANF recipients, this seems unlikely.[33] To confirm this, in results not presented here but available upon request, I plot refugee arrivals for each state by year against TANF maximum benefit levels, and find that in 26 states, they are positively correlated, and in 21, they are negatively correlated.

It is still possible that refugee populations as they were in 1996 predict the evolution of the generosity of cash assistance. To investigate this, in results not presented here I look at the relationship between total arrivals between 1983 and 1995 and evolution of cash assistance between 1996 and 2015, finding no significant association. It appears unlikely, therefore, that states adjust their cash assistance in response to the size of their refugee populations. It is important to note, however, that states would have to adjust their cash assistance in response to certain characteristics of refugee populations in order to bias my results.[34]

[33] To my knowledge, exact statistics on the number of refugees enrolled in TANF are not available. For a quick approximation, note that according to the 2015 ORR Annual Survey of Refugees, 18 percent of refugee households that had arrived between 2011 and 2015 used TANF. According to the ORR, between 2011 and 2015 approximately 634,000 refugees, Cuban/Haitian entrants, and other people eligible for ORR services arrived in the U.S. (this does not count unaccompanied minors). Translating these two facts into an estimate of the number of refugees on TANF is difficult, since we don't know the size composition of the refugee households. Assuming 18 percent of refugee individuals are on TANF gives 114,124 TANF enrollees, which is just 3.6 percent of total TANF enrollment in 2015.

[34] Given that many rules adjust their benefit levels over time according to stated rules, in Table A3, I show the results of instrumental variables specifications that use these rules as an instrument for actual benefit levels. These specifications drop states that do not use rules to adjust benefit levels. The results are similar.

1.8 Discussion

This paper explores the effect of an additional $100 in maximum benefits on refugee outcomes, finding significant effects on wages, but no significant effect on employment. The overall effect on wages was 7.7 percent in my preferred specification, which is intent-to-treat. Based on Yearbook of Immigration Statistics numbers, about 81 percent of my sample are likely to be refugees.[35] This gives an estimate of the true average treatment effect of 9.5 percent.[36]

The median real yearly wage income for working refugees interviewed between 2012 and 2016 was $20,436, so this amounts to $1,941 in additional wages for employed refugees per year, using an annual effect size of 9.5 percent. To get a sense of the total size of the impact of an additional $100 in maximum benefits, note that the average employed refugee arrived at age 32. If we assume that the effect begins five years after arrival and lasts until retirement at age 65, then we get 33 years per refugee on average of $1,941 extra wages. Using a discount rate of 5 percent, this means an average total increase in wages of $31,676 per employed refugee.[37]

On the other hand, looking at the cost of benefit provision, the cost is $100 per month to achieve this treatment effect. These data do not have information on when and how long the refugees use cash assistance, but the 2015 Annual Survey of Refugees suggests that use

[35] To calculate this, I used Table A2 to calculate the number of likely refugees in my sample from each country based on the average refugee share in column 1 and the sample size in column 3. I then added the number of likely refugees together and calculated the proportion of likely refugees in the entire sample.

[36] The 95 percent confidence interval of the effect size adjusted for the imputation is 3.0-16.0 percent.

[37] Following James J. Heckman, Seong Hyeok Moon, Rodrigo Pinto, Peter A. Savalyev and Adam Yavitz (2010) I conduct sensitivity analysis with different discount rates. No discounting leads to an estimate of $64,053, while a very conservative discount rate of 20 percent leads to an estimate of $9,699, still higher than conservative estimates of the costs of provision.

of any kind of cash assistance fell nearly in half from 72.4 percent to 39.9 percent within a year of arrival (ORR, 2016). Assuming the average refugee is on cash assistance for a year, the benefits can be compared with a cost of $1,200. Very conservatively assuming that every employed refugee remains on cash assistance for five years leads to a total cost of $6,000. This means that the increased wages that employed refugees earn are greater, on average, than the cost of providing the benefits.[38]

Only around 64 percent of refugees in my sample are employed, meaning that the costs should be scaled up to account for the refugees that use cash assistance and do not get wage benefits. Incorporating the costs of cash benefits for non-working refugees (assuming both working and non-working refugees stay on cash assistance for five years) leads to a total net benefit of $14,273 per refugee, on average.[39]

Of course, comparing the costs and benefits in this manner assumes that the benefit to the refugee of increased wages is the only outcome of interest: a full cost-benefit analysis would calculate the fiscal impacts as well as the benefit to society as a whole of the improved assimilation outcomes. However, additional cash benefits to refugees appear to be a very cost effective intervention for improving long-term outcomes.

1.9 Conclusion

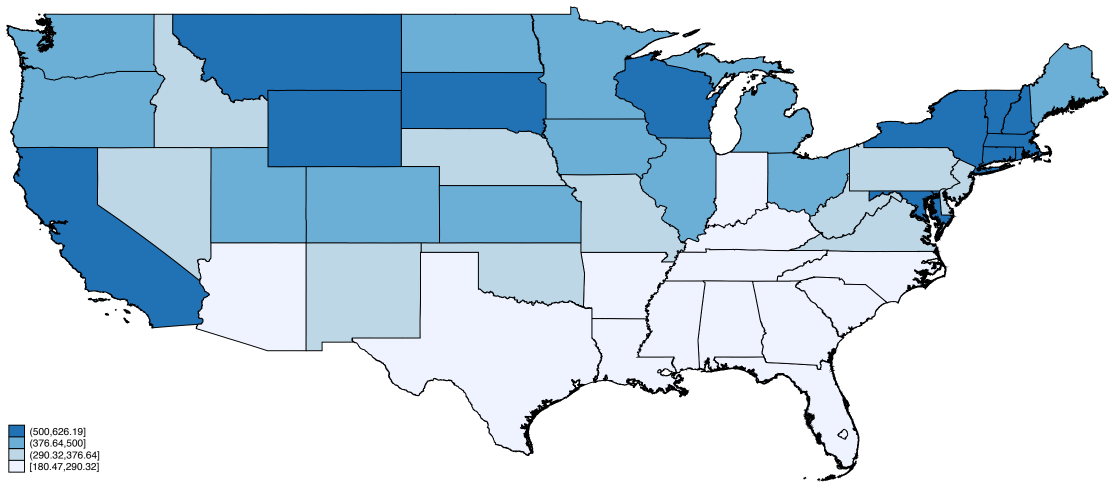
The recent conflict in Syria has brought worldwide refugee levels to post-war highs, bringing refugee resettlement to the forefront of policy debate. Given the significant public investment in refugee resettlement, the question of the long-term cost-effectiveness of cur-

[38]In this exercise, this remains true so long as the effects last at least 4 years.

[39]This is calculated as $totalbenefit = 0.64 * (31676 - 6000) + 0.36 * (-6000)$.

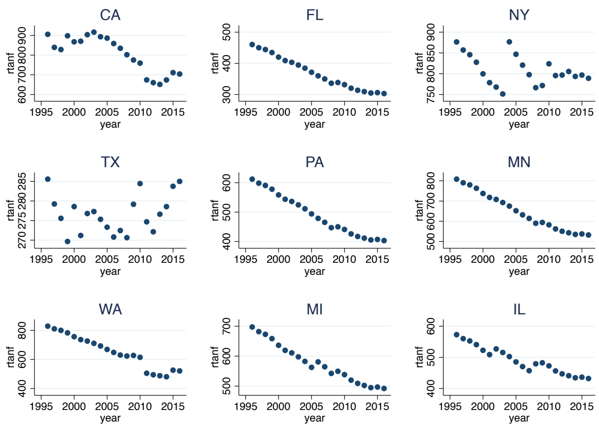
rent assistance for recently resettled refugees is of particular interest. The results of this paper suggest that the availability of larger cash assistance benefits upon arrival to the U.S. helps refugees integrate into the labor market. I find that refugees with access to larger cash benefits earn higher wages, with little to no change in probability of employment. The magnitudes of the effects imply that increased benefits are a cost-effective intervention for improving assimilation outcomes for refugees. Given the overall complexity of welfare offerings as well as the large number of degrees of freedom that states have in designing TANF and other programs, the results should be taken as suggestive. However, these results suggest that refugees, in particular highly-skilled refugees, may benefit from additional liquidity in the years directly following placement while they transfer their skills to the U.S. context and find a suitable job match.

Figure 1.1: Identifying Variation: Maximum Monthly Benefit by State



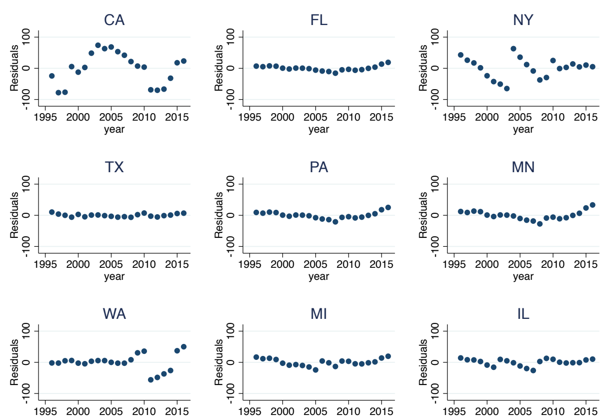
Note: Map displays TANF maximum monthly benefit for a family of three in 2015, after adjusting for price level differences using the Bureau of Economic Analysis Implicit Regional Price Deflator.
Source: Bureau of Economic Analysis; Welfare Rules Database; Floyd and Schott (2015).

Figure 1.2: Identifying Variation: Maximum Monthly Benefit by Year



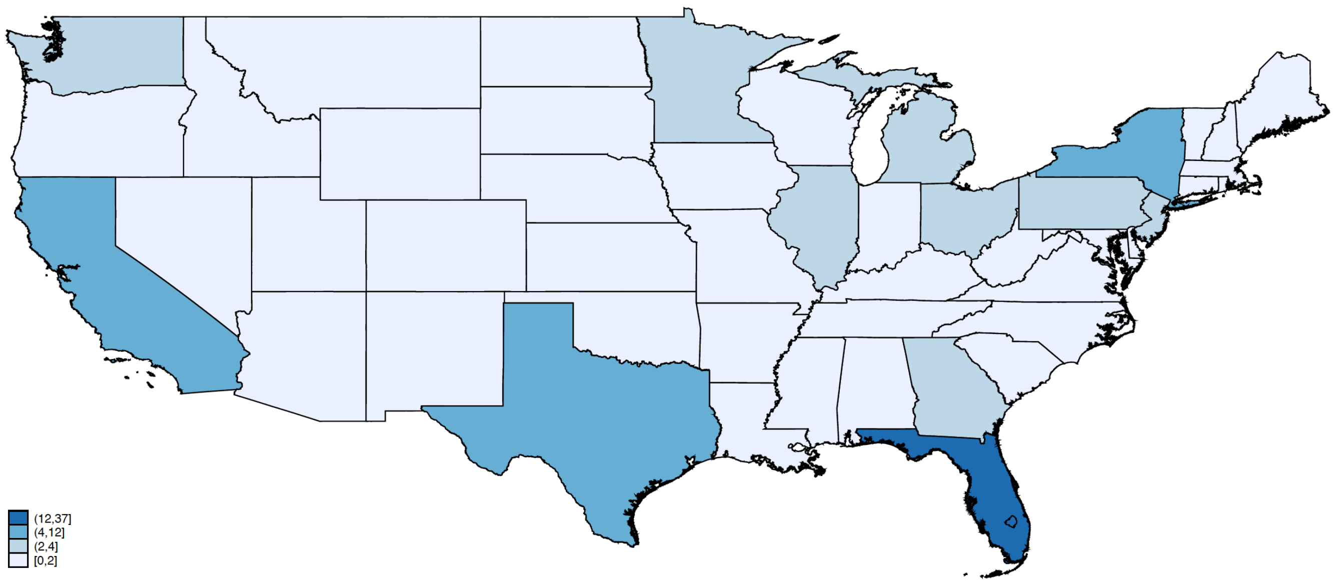
Note: Chart displays real maximum monthly benefits for a family of three by state for the nine largest resettlement states in the full dataset. The axes differ by state to better show sources of variation.
Source: Welfare Rules Database; Floyd and Schott (2015).

Figure 1.3: Maximum Monthly Benefit for a Family of 3 by Year with Linear Trends Removed



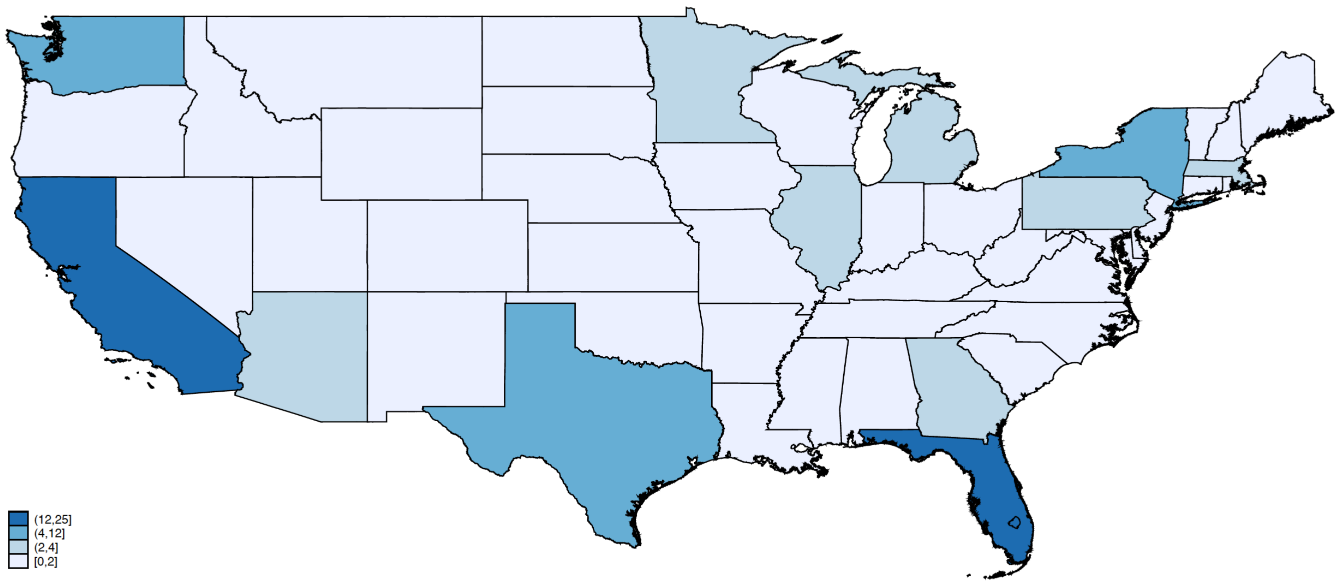
Note: Chart displays real maximum monthly benefits for a family of three by state for the nine largest resettlement states in the full Census and ACS panel dataset. Linear trends have been removed for each state; each figure shows the residuals of a regression of benefit levels on year. Therefore, the remaining variation is shown as deviations from the horizontal line at 0. The plots have all been graphed on the same y-axis scale in this figure.
Source: Welfare Rules Database; Floyd and Schott (2015).

Figure 1.4: Percent of Refugee Population by State



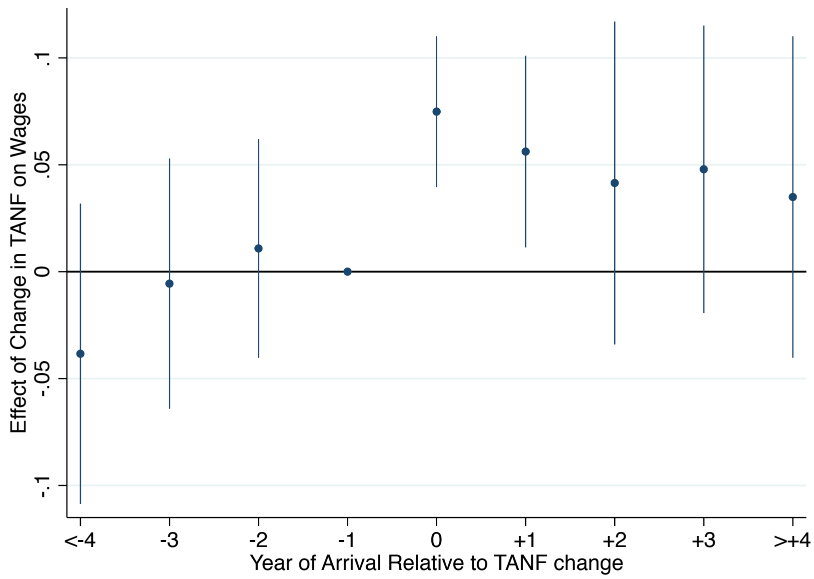
Note: Figure displays the percent of all refugees in the sample residing in each state.
Source: 2000 Census, 2001-2015 American Community Survey.

Figure 1.5: Percent of Refugee Population by State of Initial Resettlement



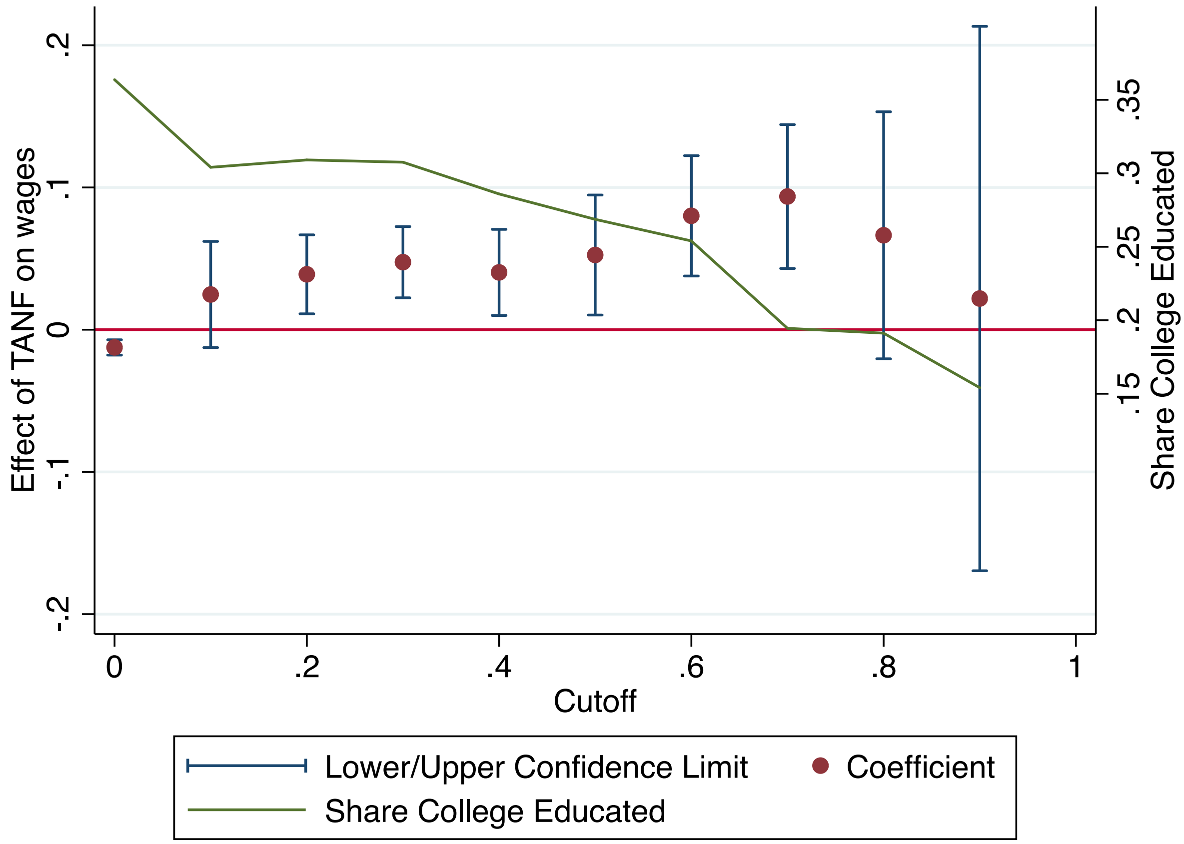
Note: Figure displays the percent of all refugees that were resettled in each state between 1996 and 2015.
Source: Office of Refugee Resettlement; author's calculations.

Figure 1.6: Event Study: Refugee Wages Before and After Large TANF Increase



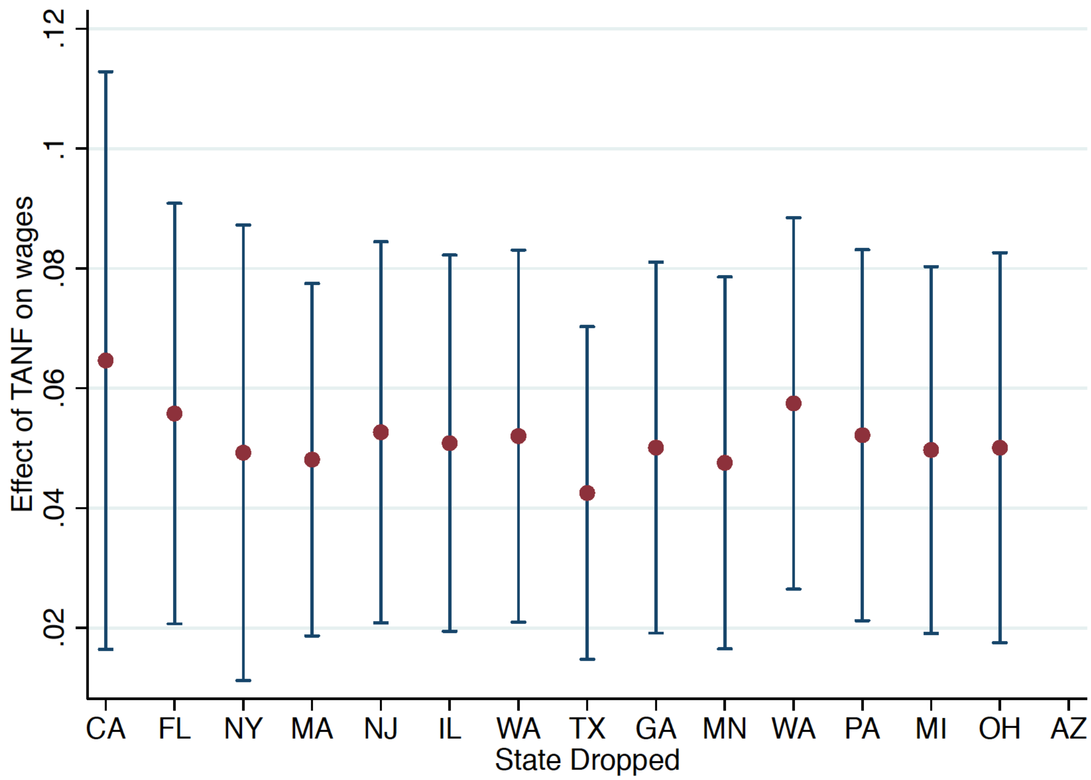
Note: This figure displays the results of a version of 1.1, where the independent variables of interest are indicator variables for whether a refugee arrived in the year of a large increase in TANF maximum benefit levels (at least 3 percent) as well as whether they arrived up to 4 years before or after such an increase. The dependent variable is log wages of employed refugees. The categories of "4+" and "-4" also include refugees who arrived greater than 4 years away from the increase in either direction. Refugees in states that never experienced a large increase in TANF benefits are omitted in this sample. The regression also includes controls for marital status, sex, age and its square, number of own children in the household, unemployment rate in the year of arrival and the year of interview, and dummy variables for each of 14 sending countries, defined as in Table A2. There are also controls for year of arrival, year of interview, and state fixed effects as well as state time trends. The standard errors are clustered at the state level.

Figure 1.7: Robustness Check: Effects of TANF Generosity with Alternative Refugee Sample Definitions



Note: This figure contains results from Equation 1.1 using various alternative criteria for identifying likely refugees. Each red circle represents the point estimate on the measure of TANF generosity using a different cutoff to define probable refugees, while the blue bars represent 90 percent confidence intervals. From the left, the first estimate gives results where all foreign born individuals are included in the sample. The second estimate uses a different definition of a probable refugee: a foreign-born individual that arrived in a year that his or her country of origin sent refugees as a fraction of refugees and economic immigrants exceeding 0.1. Moving to the right, the cutoffs increase by 0.1. On the second y axis, the relationship between the sample used and the fraction of the sample with college education is plotted. The figure shows a potential correlation between the size of the positive effect and the fraction of the sample that is college educated, which is not surprising given that college educated refugees see the largest effects of TANF generosity. All regressions exclude refugees with 0 earnings. The regressions also include controls for marital status, sex, age and its square, number of own children in the household, unemployment rate in the year of arrival and the year of interview, and dummy variables for each of 14 sending countries, defined as in Table A2. All specifications control for year of arrival, year of interview, and state fixed effects as well as state time trends. All standard errors are clustered at the state level.

Figure 1.8: Robustness Check: Effect of TANF Generosity on Wages, Dropping Top States from Sample



Note: This figure contains results from my main estimating equation on the sample of refugees arriving since 1996, dropping each of the 15 states with the largest refugee populations individually. The red circles represent point estimates, and the blue bars show 90 percent confidence intervals. The regressions exclude refugees with 0 earnings. The regressions also include controls for marital status, sex, age and its square, number of own children in the household, unemployment rate in the year of arrival and the year of interview, and dummy variables for each of 14 sending countries. The regressions include year of arrival fixed effects, year of interview and state fixed effects, and state time trends. All standard errors are clustered at the state level.

Table 1.1: Summary Statistics: Refugees 25 and Over by Region

	Soviet Union	Yugoslavia	Southeast Asia	Middle East	Cuba	Africa
Age	49.443	44.521	40.514	42.058	46.854	40.04
Marital Status	0.658	0.711	0.791	0.689	0.509	0.554
Share Female	0.568	0.499	0.495	0.478	0.509	0.503
Real Annual Wage	37,966.468	30,475.114	14,734.741	15,074.886	17,492.129	18,788.522
Number of own children	0.91	1.084	1.723	1.519	0.774	1.764
Years in the United States	17.04	15.246	4.203	5.649	8.718	9.828
English Ability	0.738	0.768	0.347	0.628	0.383	0.743
Employed	0.668	0.757	0.593	0.474	0.653	0.658
Labor Force Status	0.704	0.795	0.637	0.562	0.711	0.741
Urban Status	0.953	0.968	0.942	0.977	0.974	0.948
Enrolled in School	0.053	0.047	0.046	0.115	0.047	0.146
Years of Education	14.938	12.299	7.283	12.045	12.377	10.879
Migrated between States in Prior Year	0.014	0.007	0.04	0.028	0.015	0.044
Observations	4,864	2,905	3,285	3,471	16,860	3,182

Source: 2000 Census; 2012-2016 American Community Survey. A list of sending countries that comprise each sending region and their respective sample sizes is available in Table A2.
Note: This table gives the means of major control variables by sending region for the sample of refugees aged 25 and older.

Table 1.2: Balance Table: Refugee Characteristics Largely Do Not Predict TANF Maximum Benefit Levels

	TANF Maximum Benefit	
Dependent Variable	(1)	(2)
Marital Status	-0.010	0.007
	(0.010)	(0.010)
Female	0.018	0.009
	(0.015)	(0.009)
Age	-0.311	-0.150
	(0.765)	(0.443)
Number of Own Children	0.053	0.069**
	(0.048)	(0.029)
< High School	0.023*	0.017*
	(0.011)	(0.010)
High School	-0.017	-0.021*
	(0.026)	(0.011)
Some College	-0.002	-0.006
	(0.009)	(0.006)
College and Above	-0.004	0.010
	(0.027)	(0.011)
English Ability	-0.011	-0.005
	(0.019)	(0.008)
Soviet Union	-0.103	-0.007
	(0.110)	(0.023)
Yugoslavia	-0.008	0.012
	(0.062)	(0.016)
Southeast Asia	0.050***	0.001
	(0.015)	(0.020)
Middle East	0.007	-0.019**
	(0.026)	(0.008)
Cuba	0.020	0.002
	(0.032)	(0.010)
Africa	0.035	0.012
	(0.041)	(0.033)
Observations	78,129	104,227
Sample	>60%	>40%

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Note: This table contains results from Equation 1.1, run without controls, on the sample of refugees arriving since 1996. The outcome variables are refugee characteristics. In column 1, refugees are defined as individuals from country-years where refugees made up at least 60 percent of new arrivals from their country of origin. In column 2, I use a cutoff of 40 percent to expand the sample. All specifications control for year of arrival, year of interview, and state fixed effects, as well as state time trends. All standard errors are clustered at the state level.

Table 1.3: Main Results: Effects of TANF Generosity on Wages

Dependent variable:	Log Wage	Log Wage	Log Wage	Log Wage
	(4)	(4)	(4)	(4)
TANF Max Benefit	0.029	0.045	0.052***	0.077***
	(0.029)	(0.036)	(0.019)	(0.026)
< High School			-0.392***	-0.468***
			(0.025)	(0.033)
High School			-0.334***	-0.389***
			(0.030)	(0.032)
Some College			-0.291***	-0.333***
			(0.050)	(0.052)
English Ability			0.233***	0.220***
			(0.015)	(0.014)
Observations	50,878	34,479	50,878	34,479
Controls	No	No	Yes	Yes
Sample	All	In U.S. >4 Years	All	In U.S. >4 Years

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Note: This table contains results from Equation 1.1 on the sample of refugees arriving since 1996, excluding those with 0 earnings. The regressions in columns 3 and 4 also include controls for marital status, sex, age and its square, number of own children in the household, unemployment rate in the year of arrival and the year of interview, and dummy variables for each of 14 sending countries, defined as in Table 1.1. Columns 2 and 4 exclude refugees who have been in the U.S. for fewer than five years, as they may still receive TANF benefits. All specifications control for year of arrival, year of interview, and state fixed effects, as well as state time trends. All standard errors are clustered at the state level.

Table 1.4: Effects of TANF Generosity on Wages for Men and Women

Dependent variable:	Log Wage			
	(1)	(2)	(3)	(4)
TANF Max Benefit	0.038*	0.117**	0.079***	0.033
	(0.021)	(0.055)	(0.027)	(0.033)
< High School	-0.378***	-0.456***	-0.405***	-0.475***
	(0.038)	(0.042)	(0.030)	(0.030)
High School	-0.329***	-0.373***	-0.339***	-0.411***
	(0.039)	(0.039)	(0.025)	(0.029)
Some College	-0.293***	-0.332***	-0.288***	-0.336***
	(0.059)	(0.054)	(0.041)	(0.050)
English Ability	0.226***	0.196***	0.244***	0.250***
	(0.018)	(0.013)	(0.029)	(0.034)
Observations	28,074	18,426	22,804	16,053
Sex	Men	Men	Women	Women
Sample	All	In U.S. >4 Years	All	In U.S. >4 Years

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Note: This table contains results from Equation 1.1 on the sample of refugees arriving since 1996 run separately on men (columns 1 and 2) and women (columns 3 and 4), excluding those with 0 earnings. The regressions also include controls for marital status, sex, age and its square, number of own children in the household, unemployment rate in the year of arrival and the year of interview, and dummy variables for each of 14 sending countries, defined as in Table 1.1. Columns 2 and 4 exclude refugees who have been in the U.S. for fewer than five years, as they may still receive TANF benefits. All specifications control for year of arrival, year of interview, and state fixed effects, as well as state time trends. All standard errors are clustered at the state level.

Table 1.5: Main Results by Education: Interaction Between TANF Generosity and Education

Dependent Variable:	Log Wage			
		(2)	(3)	(4)
TANF Max Benefit	0.083***	0.105***	0.097***	0.110***
	(0.016)	(0.022)	(0.021)	(0.029)
< High School	-0.132***	-0.193***	-0.117**	-0.192***
	(0.044)	(0.054)	(0.049)	(0.071)
High School	-0.113*	-0.166***	-0.12	-0.168**
	(0.061)	(0.059)	(0.077)	(0.074)
Some College	-0.039	-0.106	-0.078	-0.187*
	(0.077)	(0.098)	(0.072)	(0.097)
TANF*<HS	-0.050***	-0.052***	-0.047***	-0.046***
	(0.009)	(0.011)	(0.011)	(0.015)
TANF*HS	-0.042***	-0.042***	-0.037**	-0.037**
	(0.014)	(0.011)	(0.018)	(0.015)
TANF*Some College	-0.046***	-0.041***	-0.033**	-0.02
	(0.013)	(0.014)	(0.014)	(0.015)
English Ability	0.232***	0.221***	0.239***	0.222***
	(0.015)	(0.014)	(0.016)	(0.015)
Observations	50,878	34,479	39,621	24,788
Controls	Yes	Yes	Yes	Yes
Sample	All	In U.S. >4 Years	All	In U.S. >4 Years
Age	Full Sample	Full Sample	Over 24	Over 24

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Note: This table contains results from Equation 1.1 on the sample of refugees arriving since 1996, allowing for interaction effects between TANF generosity and each of the educational categories. Columns and 2 include the full sample of refugees above the age of 24, while columns 3 and 4 exclude those arriving prior to the age of 25. All regressions exclude refugees with 0 earnings. The regressions also include controls for marital status, sex, age and its square, number of own children in the household, unemployment rate in the year of arrival and the year of interview, and dummy variables for each of 14 sending countries, defined as in Table A2. Columns 2 and 4 exclude refugees who have been in the U.S. for fewer than five years, as they may still receive TANF benefits. All specifications control for year of arrival, year of interview, and state fixed effects, as well as state time trends. All standard errors are clustered at the state level.

Table 1.6: Mechanisms: Effects of TANF Generosity on Occupational Characteristics

Dependent variable:	Occ 1990 Avg Educ (1)	Occ 2000 Avg Educ (2)	Occ 1990 Avg Wage (3)	Occ 2000 Avg Wage (4)
TANF Max Benefit	0.176***	0.141**	0.025*	0.019
	(0.048)	(0.058)	(0.013)	(0.015)
< High School	-2.008***	-2.061***	-0.343***	-0.357***
	(0.111)	(0.106)	(0.030)	(0.028)
High School	-1.821***	-1.867***	-0.321***	-0.340***
	(0.144)	(0.145)	(0.031)	(0.031)
Some College	-1.335***	-1.389***	-0.257***	-0.274***
	(0.127)	(0.129)	(0.031)	(0.030)
English Ability	0.635***	0.646***	0.104***	0.109***
	(0.031)	(0.034)	(0.008)	(0.008)
Observations	34,479	34,479	34,479	34,479

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Note: This table contains results from Equation 1.1 on the sample of refugees arriving since 1996, excluding those with 0 earnings. The regressions in columns 1 and 2 use occupational mean education in years as the outcome, while columns 3 and 4 use occupational mean wage in 1990 and 2000, respectively, as the outcome variable. All regressions include controls for marital status, sex, age and its square, number of own children in the household, unemployment rate in the year of arrival and the year of interview, and dummy variables for each of 14 sending countries, defined as in Table A2. All regressions exclude refugees who have been in the U.S. for fewer than five years, as they may still receive TANF benefits. All specifications control for year of arrival, year of interview, and state fixed effects, as well as state time trends. All standard errors are clustered at the state level.

Table 1.7: Mechanisms: Effects of TANF Generosity on Alternative Outcomes

	Employed	Food Stamp Use	Moved States	English Ability	Employed	Food Stamp Use	Moved States	English Ability
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
TANF Max Benefit	0.008	0.001	0.006*	-0.01	0	-0.012	-0.001	-0.019
	(0.017)	(0.010)	(0.003)	(0.018)	(0.015)	(0.012)	(0.004)	(0.037)
< High School	-0.164***	0.137***	-0.010***	-0.371***	-0.194***	0.158***	-0.009***	-0.380***
	(0.013)	(0.011)	(0.004)	(0.029)	(0.007)	(0.012)	(0.003)	(0.017)
High School	-0.068***	0.062***	-0.008***	-0.238***	-0.083***	0.070***	-0.008***	-0.241***
	(0.006)	(0.006)	(0.002)	(0.010)	(0.006)	(0.005)	(0.002)	(0.022)
Some College	-0.038***	0.042***	-0.008***	-0.100***	-0.051***	0.050***	-0.007***	-0.098***
	(0.007)	(0.006)	(0.002)	(0.009)	(0.007)	(0.005)	(0.002)	(0.007)
English Ability	0.098***	-0.103***	0.002**		0.098***	-0.101***	0.002	
	(0.017)	(0.011)	(0.001)		(0.019)	(0.013)	(0.001)	
Observations	78,129	67,977	67,977	78,129	51,072	51,072	51,072	51,072
Sample	All	All	All	All	In U.S. >4 Years	In U.S. >4 Years	In U.S. >4 Years	In U.S. >4 Years

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Note: This table contains results from Equation 1.1 on the sample of refugees arriving since 1996, using various alternative outcome variables. The regressions in the first four columns are run on the entire sample, including refugees with 0 earnings, while those in columns 5-8 also exclude refugees who have been in the U.S. for fewer than five years. The regressions also include controls for marital status, age and its square, number of own children in the household, unemployment rate in the year of arrival and the year of interview, and dummy variables for each of 14 sending countries, defined as in Table A2. The specifications control for year of arrival, year of interview, and state fixed effects, as well as state time trends. All standard errors are clustered at the state level.

Table 1.8: Heterogeneity in Employment Effects of Cash Generosity

Dependent Variable:	Employment		
	(1)	(2)	(4)
TANF Max Benefit	0.006	-0.013	-0.005
	(0.014)	(0.017)	(0.014)
< High School	-0.143***	-0.195***	-0.194***
	(0.018)	(0.007)	(0.007)
High School	-0.029**	-0.082***	-0.083***
	(0.013)	(0.005)	(0.006)
Some College	-0.018	-0.049***	-0.051***
	(0.013)	(0.006)	(0.007)
TANF*<HS	-0.009***		
	(0.002)		
TANF*HS	-0.010***		
	(0.003)		
TANF*Some College	-0.006**		
	(0.003)		
English Ability	0.098***	-0.011	0.098***
	(0.019)	(0.014)	(0.019)
TANF*English		0.021***	
		(0.002)	
Female	-0.148***	-0.149***	-0.190***
	(0.006)	(0.006)	(0.014)
TANF*Female			0.008**
			(0.003)
Observations	51,072	51,072	51,072
Interaction	Education	English Ability	Female

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Note: This table contains results from Equation 1.1 on the sample of refugees arriving since 1996, using employment as the outcome variable of interest. Each column shows the results with the generosity measure interacted with a refugee characteristic: column 1 shows interactions with the education categories, column 2 shows interactions with English ability, and column 3 shows interactions with sex. The regressions include controls for marital status, sex, age and its square, number of own children in the household, unemployment rate in the year of arrival and the year of interview, and dummy variables for each of 14 sending countries, defined as in Table A2. All specifications control for year of arrival, year of interview, and state fixed effects, as well as state time trends. All standard errors are clustered at the state level.

Table 1.9: Robustness Checks: Effects of TANF Generosity with Sample Exclusions

Dependent Variable:	Log Wage		
	(1)	(2)	(3)
TANF Max Benefit	0.054*	0.055	0.071**
	(0.031)	(0.036)	(0.027)
< High School	-0.469***	-0.524***	-0.470***
	(0.038)	(0.034)	(0.034)
High School	-0.391***	-0.447***	-0.391***
	(0.037)	(0.037)	(0.033)
Some College	-0.324***	-0.413***	-0.335***
	(0.058)	(0.029)	(0.054)
English Ability	0.220***	0.194***	0.220***
	(0.015)	(0.030)	(0.014)
Observations	29,161	19,721	33,583
Publicly Administered Only	Yes	No	No
Excl. Cubans	No	Yes	No
Excl. States with Time Limit Changes	No	No	Yes

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Note: This table contains results from Equation 1.1 on various subsamples of refugees arriving since 1996. Column 1 excludes refugees not resettled in a state using a publicly administered program for RCA. Column 2 excludes Cubans. Column 3 excludes refugees resettled in states that experienced a change in TANF time limit policy during the sample period. All regressions exclude refugees with 0 earnings. The regressions also include controls for marital status, sex, age and its square, number of own children in the household, unemployment rate in the year of arrival and the year of interview, and dummy variables for each of 14 sending countries, defined as in Table A2. All specifications control for year of arrival, year of interview, and state fixed effects as well as state time trends. All standard errors are clustered at the state level.

Table 1.10: Effects of TANF Generosity on Non-Refugee Immigrants

Dependent Variable:	Log Wage			
	(1)	(2)	(3)	(4)
TANF Max Benefit	-0.027***	-0.022***	-0.009**	-0.011***
	(0.009)	(0.008)	(0.004)	(0.003)
< High School			-0.700***	-0.733***
			(0.024)	(0.022)
High School			-0.605***	-0.646***
			(0.020)	(0.018)
Some College			-0.511***	-0.541***
			(0.020)	(0.019)
English Ability			0.260***	0.246***
			(0.011)	(0.011)
Observations	945,297	656,011	945,297	656,011
Controls	No	No	Yes	Yes
Sample	All	In U.S. >4 Years	All	In U.S. >4 Years

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Note: This table contains results from Equation 1.1 on the sample of non-refugee foreign born arriving since 1996, excluding those with 0 earnings. The regressions in columns 3 and 4 also include controls for marital status, sex, age and its square, number of own children in the household, unemployment rate in the year of arrival and the year of interview, and dummy variables for each of 11 sending regions. These regions are Mexico, Canada, Latin America (excluding Mexico), Northern and Western Europe, Eastern Europe, East Asia, Southeast Asia, Southwest Asia, the Middle East, Africa, and Oceania. Columns 2 and 4 exclude refugees who have been in the U.S. for fewer than five years. All specifications control for year of arrival, year of interview, and state fixed effects, as well as state time trends. All standard errors are clustered at the state level.

Chapter 2

Temperature, Worker Productivity, and Adaptation: Evidence from Survey Data Production

2.1 Introduction

How environmental factors impact economic outcomes is an open question that has drawn significant interest in the economics literature (Melissa Dell, Benjamin F. Jones and Benjamin A. Olken, 2014). While a growing body of work documents negative effects of high temperature on aggregate outcome variables such as GDP and labor income,[1] less is known about the role of temperature in individual worker productivity and behavior. Understanding the margins of adjustment for productivity impacts on hot days and how they fit in with incentives may uncover welfare implications of temperature that are not captured in aggregate productivity measures (Geoffrey Heal and Jisung Park, 2016). However, few datasets are well suited for these analyses, especially in climate-vulnerable developing countries.

This paper overcomes data constraints by using household survey datasets to examine the behavior of the interviewer rather than the respondent. I examine how temperature and humidity on the day of interview impact the collection of household survey data by Demographic and Health Survey (DHS) interviewers from 46 developing countries. In developing

[1]See, for example: Melissa Dell, Benjamin F. Jones and Benjamin A. Olken (2012), Solomon M. Hsiang (2010), Marshall Burke, Solomon M. Hsiang and Edward Miguel (2015), Geoffrey Heal and Jisung Park (2014), Tatyana Deryugina and Solomon M. Hsiang (2014).

countries, interviewers' jobs require them to walk significant distances between households and conduct interviews face-to-face in non climate-controlled settings, making them members of the group of workers in poor countries significantly exposed to outdoor temperatures. Each interview in the DHS contains the interviewer's unique identifier, a start time and an end time in minutes, and a host of rich information on the quality of data collected in the field, allowing me to reconstruct the daily work schedules of over 9,000 interviewers.

Temperature has well-known negative impacts on human physiology. Experiments in laboratory and single-firm settings have found that heat has significant impacts on both physical and cognitive task performance (see Olli Seppanen, William J Fisk and QH Lei (2006), for example) and may cause a "slowing-down" of work (Tord Kjellstrom, Ingvar Holmer and Bruno Lemke, 2009). Temperature may affect the disutility of effort, which is potentially important in principal-agent scenarios with limited monitoring. The rich data on the timing of interviewer activities, the quantity and quality of interviewer output, and the workers' incentives make it possible to study the margins of adjustment on hot days and how workers minimize utility loss from high temperatures.

The DHS provides a useful setting for examining the impacts of temperature on productivity due to its enormous spatial and temporal coverage, but also due to the fact that the process for hiring interviewers, conducting interviews, and processing data is consistent across contexts. The DHS produces and publishes regular guidelines for field supervisors and editors for both managing fieldwork and evaluating interviewer performance. Interviewers are paid a fixed amount per day in most cases rather than a per hour or piece-rate wage. The documentation suggests that it is the number of completed interviews that is the measure of productivity most observable to the interviewers' supervision teams, while data quality is arguably more difficult for the supervisors to observe. I exploit the multidimensional nature of interviewer jobs combined with differential ease of monitoring to test the prediction that worker productivity responds especially strongly to temperature fluctuations on dimensions with less intensive monitoring. Furthermore, the fact that the dataset contains dozens of countries allows me to compare standardized measures of worker productivity across institutional contexts and examine differences in the impacts of temperature across settings.

Using a semi-parametric specification to allow for nonlinearities in the effects of temperature, I find that the total quantity of production is insensitive to temperature, but that the pace of production slows by approximately 20 percent in the hottest and most humid days. The results suggest that interviewers work longer days in order to achieve the same daily output of interviews, potentially because the incentives to do so are strong due to supervisor monitoring of that task. I use wet bulb temperature, a heat index that combines temperature and humidity, as the independent variable of interest, and I identify impacts off of variation in weather within a region of a country and accounting for average local weather in the month of interview to isolate the causal effect of temperature.

I find that data quality problems, such as missing responses and flags for data quality added in the data processing stage, become more frequent on hot days. Interviews contain 0.3-0.4 more missing responses on hot days, a large increase relative to an average of one total per interview. The fact that the number of completed interviews does not decrease while data quality suffers suggests that workers reallocate their effort to sustain productivity on more observable dimensions. I also find suggestive evidence that impacts on productivity vary significantly with the quality of public sector management in the implementing country. Quality of public sector management may proxy for overall quality of implementation or monitoring, since most surveys are implemented by public agencies in the country being surveyed. In countries with poor public sector management, as measured by the World Bank Country Policy and Institutional Assessment, interviewers complete fewer interviews on hot days and work fewer hours, implying that the impacts of temperature vary significantly by institutional setting and that worse overall management may weaken incentives around number of completed interviews.

I find that interviewers' days start earlier as average temperature increases, both in the cross section and using the regression framework. However, the association is smaller once region fixed effects and local climate controls are accounted for (an effect of 30 minutes as opposed to two hours). This suggests that workers in places that are warmer on average start their days earlier on average, not just on days that are surprisingly hot, providing evidence that some forms of adaptation to local climates exacerbate rather than alleviate the impacts of temperature on productivity in terms of output per time worked. I also find that workers spend more time on each interview and that more interviews start in the early morning and late afternoon hours on hot days. This suggests that interviewers allocate their work effort to hours with more pleasant temperatures. However, workers have fewer hours in a day for leisure if they are working more hours and must work longer to achieve the same output. This implies that adaptation behavior may not be uniformly costless: as climates change, individuals may adopt behaviors that, while reducing utility costs of temperature changes on net, have costs in other ways. On the other hand, I find that workers are less likely to conduct interviews at all on hot days, a form of adaptation that likely ameliorates the productivity effects of high temperature.

The interviewers' jobs require them to interact productively with another person who is also potentially exposed to the same variation in temperature: the respondent. It is therefore possible the interviewer's job becomes more difficult on hot days. The main result of this paper treats the impact of temperature on number of interviews per hour as inclusive of effects on the respondent and the interviewer; however, this paper aims to give a picture of the impacts of temperature on interviewer effort. Therefore, in Section 2.6 I discuss in depth the evidence on the role of the respondent, finding suggestive evidence that respondents do not drive the main results.[2]

While data production is a unique context due, in part, to the role of the respondent, anecdotal evidence from the developing world suggests that lessons learned from household survey interviewers may generalize to other sectors.[3] Due to factors such as a lack of indoor climate control and poverty, heat exposure in the workplace in developing countries is not restricted to primary sector occupations such as agriculture: service sector workers in areas such as community services, transportation, and tourism are significantly affected (Tord Kjellstrom, David Briggs, Chris Freyberg, Bruno Lemke, Matthias Otto and Olivia Hyatt, 2016*a*). Furthermore, a recent report suggests that many workers in the developing world are subject to daily production targets and non-climate controlled workplaces, which may yield similar results (Tord Kjellstrom, Matthias Otto, Bruno Lemke, Olivia Hyatt, Dave Briggs, Chris Freyberg and Lauren Lines, 2016*b*).[4] All told, the findings suggest that about 5 percent

[2] The respondent could impact the results either through an effect on nonresponse, leading to selection into the sample or to more unsuccessful interview attempts, or by behaving differently during the interview. In Section 2.6, I show that there is little evidence of selection on observable characteristics driving the main results of this paper. Furthermore, there is evidence that interviewers' incentives are an important component of the results, suggesting respondent behavior is unlikely to be driving my estimates.

[3] Furthermore, as I discuss in Section 2.3, interviewer productivity is correlated with aggregate productivity, and the measures I use evolve with experience in a manner consistent with usual measures of task productivity.

[4] The report says that "Heat stress and the same daily production targets in all parts of the year means

more interviewer hours in the field were required under actual weather distributions than would have been required if each interviewing day had a mild average temperature.

This paper contributes to the previous literature on temperature, productivity, and worker effort along several dimensions. First, this paper ties into the literature on temperature and economic outcomes (see Heal and Park (2016) for a review) by estimating a reduced-form impact of temperature on the productivity of outdoor workers in 46 countries. This provides the broadest estimate yet of the relationship between temperature and productivity at the individual level, particularly in climate-vulnerable developing countries. This estimate focuses on a set of workers whose job requires significant exposure to the outdoors in addition to successful interaction with respondents, making them a significantly different sample of workers from those in a manufacturing job, as have been previously studied in developing country contexts.

Second, the paper examines margins of adjustment on hot days, adding to the growing literature on temperature and adaptive behavior in economics.[5] Graff Zivin and Neidell (2014) find that workers in climate-exposed industries in the U.S. work fewer hours on hot days and that individuals spend less time on outdoor leisure. That paper conjectures that the impacts of temperature on labor supply may be larger in developing countries, where more workers are exposed to outdoor temperatures, which are also higher in developing

that the workers have to work longer each day in the hot season than in cool seasons, but the salaries typically remain the same"

[5]Typically, as Joshua Graff Zivin and Matthew Neidell (2014) note, the literature has used the word "adaptation" to describe the long-term process by which regions with warmer climates are able to attenuate the negative effects of temperature on economic outcomes through the adoption of technology such as air conditioning or through physical acclimation. Their evidence suggests that individuals may adopt short-term behavior to avoid some of the utility losses of extreme temperature as well.

countries on average. I examine impacts on hours worked directly and find the opposite response in this setting. Heal and Park (2016) predict that working hours decrease under high temperatures given a piece-rate wage contract; in my setting, which features a fixed wage contract, working hours instead increase, shedding light on the importance of incentives in determining the impacts of temperature. Understanding the margins of adjustment as well as how they fit in with incentives is crucial for examining the welfare impacts of these extreme temperatures, especially from the standpoint of worker utility.

Finally, the paper links to the literature on worker effort and monitoring. Standard principal agent models predict a decrease in effort if personal costs of effort increase, and task-based models predict that worker effort will be focused on tasks that are especially closely monitored or highly rewarded (Bengt Holmstrom and Paul Milgrom, 1991*a*). This paper examines whether an increase in costs of effort (caused by extreme temperature) will differentially affect tasks with different levels of monitoring, providing evidence on the relationship between disutility of effort, monitoring, and provision of effort in a real-world context.

The rest of the paper proceeds as follows. Section 2.2 provides background on temperature and economic outcomes as well as the DHS interviewers, exploring the incentives that interviewers face in the field. Section 2.3 describes the dataset, which merges data from over 9,000 DHS interviewers between 1990 and 2010 with gridded global weather information. Section 2.4 describes the empirical strategy, which compares data production outcomes within a region of a country on hot and cold days versus mild days. Section 2.5 presents the main results of the paper, describing both the main effect of temperature on data production as well as evidence on margins of adjustment. Section 2.6 discusses the potential contribution of the respondent to the observed effects on production. Section 2.7 explores heterogeneity in the main results. Section 2.8 shows robustness of the main results to other specifications of temperature and the control variables. Section 2.9 puts the results into the context of the broader literature and discusses the generalizability of these findings. Section 2.10 concludes.

2.2 Background

2.2.1 Temperature and Worker Productivity

The fact that temperature has significant effects on human activity and welfare is well established and has a strong basis in human physiology. To survive, humans must maintain a body temperature within a relatively narrow range. The body has several strategies for maintaining a healthy body temperature when exposed to hot and cold temperatures; however, the more extreme the outdoor temperature, the greater stress it puts on the body.

Experiments in laboratories and individual workplaces have found that heat has significant impacts on performance on physical and cognitive tasks (see Seppanen, Fisk and Lei (2006) for a review) and may cause a "slowing-down" of work (Kjellstrom, Holmer and Lemke, 2009). Building on a solid foundation of laboratory and workplace studies, a burgeoning literature in economics has examined the impacts of fluctuations in temperature on economically-relevant outcomes such as test scores and human capital formation (Jisung Park (2018); Joshua S. Graff Zivin, Yingquan Song, Qu Tang and Peng Zhang (2018)), mortality (Olivier Deschênes and Michael Greenstone, 2009), workplace accidents (Marcus Dillender, 2017), violence (Solomon M. Hsiang, Marshall Burke and Edward Miguel, 2013), and several others. Previous studies on productivity have found negative, nonlinear effects of high outdoor temperature on the productivity of manufacturing workers in India ((Achyuta Adhvaryu, Namrata Kala and Anant Nyshadham, 2018); Bengt Holmstrom and Paul Milgrom (1991*b*)) and on total factor productivity and output in Chinese manufacturing firms (Peng Zhang, Olivier Deschenes, Kyle C. Meng and Junjie Zhang (2018); Xiaoguang Chen and Lu Yang (2017)).[6] Literature on the effect of wages on productivity has also found that this relationship depends on environmental factors such as temperature for agricultural workers in the U.S. (Andrew Stevens, 2016). Building on the foundation of these studies, this paper extends the analysis of the impacts of temperature on workers to a broader geographical context and examines the mechanisms for the effects on workers.

In addition, recent papers emphasize the role of humidity in the impacts of high temperatures on human health. As the ambient temperature becomes hotter, the human body becomes more reliant on sweating as a mechanism for cooling itself. Humidity impairs the body's ability to cool itself through sweating, multiplying the effects of heat stress. These papers show that humidity significantly amplifies the effects of heat on mortality in the United States and in developing countries (Alan Barreca (2012); Michael Geruso and Dean Spears (2018*a*)) and, given humidity levels in developing countries, may also play a large role in occupational health and productivity (Tord Kjellstrom and Jennifer Crowe (2013); John P. Dunne, Ronald J. Stouffer and Jasmin G. John (2013)). The role of humidity has implications for the distribution of climate damages as well as for the magnitudes, as the world regions that experience hot and humid weather are distinct from those that experience hot and dry weather. With this in mind, I focus on wet bulb temperature, which accounts for

[6] See Heal and Park (2016) for an excellent review of the literature on both micro- and macroeconomic impacts.

relative humidity, as the main weather variable of interest in this analysis, following recent papers in the economics literature (Adhvaryu, Kala and Nyshadham (2018); Geruso and Spears (2018*a*)).[7] This measure is described in more detail below.

2.2.2 DHS Interviewers: Contracts and Incentives

This section describes the process of hiring DHS interviewers and conducting surveys, examining the sources of workplace incentives and describing the tasks the workers complete. The DHS program has facilitated the administration of over 300 nationally and regionally representative surveys of reproductive and health behaviors in developing countries since the 1980s. The surveys are administered by local implementing agencies with technical support from the DHS program, which is funded by the U.S. Agency for International Development. The implementing agencies are most often governmental organizations such as National Statistical Offices or the Ministry of Health, but can also be non-governmental or private. The DHS program sets standard guidelines for the implementation of the surveys: manuals are publicly available for interviewers and supervisors and editors for the processes of training field staff and sampling, among other topics. The questionnaires used in the DHS are also standard within phases of the survey, although optional modules (e.g. domestic violence, HIV, and anemia) are implemented in some places but not others. These standardized procedures are put into place to ensure comparability of the data across countries. This means that interviewers' tasks are standardized across contexts, though their employers vary.

[7] In Section 2.5 I show direct evidence of a significant interaction effect between temperature and humidity in determining worker productivity.

The focus of the surveys is on the collection of accurate reproductive histories and information on the health status of women and children in developing countries. Therefore, the majority of respondents are prime-aged women (between the ages of 15 and 49), although in many survey rounds some males are interviewed in the households as well. In the majority of the more recent DHS surveys, DHS teams have also been collecting GPS coordinates for each survey cluster (e.g. a village in a rural area or a city block in an urban area) in the sample.[8] Most of these are accurate to within 15 meters.[9]

Recruitment of fieldworkers is conducted locally by implementing agencies, but the practices for recruitment and required qualifications are standard across contexts. In most cases, interviewers are temporary employees of the survey implementing organization for the duration of the survey, although sometimes the implementing agencies use their own permanent staff.[10] Interviewers must be available to work full time for the duration of the survey, including nights and weekends, and they must have sufficient physical fitness to walk long distances and carry questionnaires as required. There is a strong preference for interviewers to interview respondents of their own gender, and they must speak at least one of the languages used for training (and thus prominent in that local context). Interviewers are recruited within a region of a country as much as possible so that the interviewer does not seem too foreign to the respondent, and there is a preference for candidates with at least a secondary education. This means that in most cases, interviewers are a more highly-educated population than the average residents of their countries.[11]

[8] Clusters, or enumeration areas, are generally chosen with probability proportional to population size from a recent census.

[9] 1% of rural clusters are displaced between 0 and 10 kilometers to protect the privacy of the respondents.

[10] This is not observable in this data sample.

[11] Information on interviewer characteristics beyond the unique identifier is not available in most rounds of

The contracts in each survey round are determined by the survey implementing agency, which has freedom in designing pay structure. However, conversations with technical staff at ICF, which implements the DHS, indicate that the standard practice in the DHS is to pay interviewers a fixed amount per day plus a per diem for food/lodging/etc. Anecdotally, piece-rate wages are problematic in this context because some interviews naturally take longer than others, so piece rates would introduce significant risk into interviewers' wages, and furthermore, piece rates incentivize quantity of production over quality. The fact that interviewers are not paid piece rate, however, means that their incentive to perform well comes from any risk of termination, or from wanting to maintain a good reputation (many interviewers have previous experience and will go on to work for other surveys). There is no direct information on how many workers are fired during the average survey, but the implementing organizations are recommended to hire 10 percent more interviewers than needed to serve as reserve interviewers to fill gaps in case of separations.[12]

---

the DHS. In a few recent surveys, the DHS has included a questionnaire on interviewer characteristics.The surveys are: Afghanistan, 2015-2016; Armenia, 2015-2016; Nepal, 2016; Zimbabwe, 2015. These surveys took place after 2010, so they cannot be linked with the weather data. However, I summarize the information on interviewer education and age in Figure B9. The data reveal that levels of education actually vary by context in the DHS; interviewers in Africa have few years of education while interviewers in Afghanistan and Nepal are quite well educated. The interviewers are very young on average, and the data also reveal that almost half had worked on a prior DHS survey and nearly 75 percent had worked on another survey of some kind.

[12]In Table B1, I examine whether interviewers are more likely to "leave" the survey round before the rest of their team if they have worse average measures of productivity. I am unable to observe whether these separations occur because the interviewer, quit, was fired, or was simply finished with his or her interviews, but I document a significant correlation between interviewer productivity and having a final interviewing date before the rest of the interviewer's team.

2.2.2.1 Interviewer Tasks and Performance Evaluation

Data are collected by interviewing teams usually comprised of a supervisor, a female field editor, several female interviewers, and one to two male interviewers if men will be included in the survey. Supervisors and field editors often share responsibilities, but supervisors in general are responsible for organizing fieldwork (arranging transportation, lodging, etc.), delegating interviews to interviewers, and conducting spot check re-interviews. The focus of field editors is on data quality: field editors observe at least one interview per day and edit completed questionnaires while in the field, sending interviewers back to correct problems if necessary.[13] Each interviewing team generally is assigned its own car.

Data collection in each household begins with a household survey, in which basic information on each member of the household is recorded along with descriptive information on the household such as building materials and water sources. In selected households, all women aged 15-49 are then eligible to be interviewed using the individual questionnaire. In many survey rounds, men are also interviewed, usually in every second or third household. Reproductive history is gathered from both men and women, but only women answer detailed questions about child health and other outcomes. In interviews with women, detailed questions are asked about all children under an age cutoff, usually five. The DHS also includes anthropometric measurements of children under five and women, such as height and weight and often country-specific biomarkers such as tests for anemia, malaria, or HIV. In cases where either an individual or an entire household is not available for interview on the first visit, DHS rules stipulate that interviewers must make at least 3 visits at separate

[13] It is not observable in the DHS data which interviews were observed by a field editor.

times/days to make every effort to interview eligible individuals. In the data, the DHS identifies the interviews that required more than one attempt, although there is no information on when the unsuccessful attempts took place. As I discuss in Section 2.6, any effect of temperature on the probability of nonresponse could affect the main results on quantity and quality of data production. I investigate this empirically below.

The schedules of DHS interviewers are set at the beginning of each day by the team's field supervisor. Guidelines indicate that the supervisors give assignments in the form of lists of households that each interviewer is meant to visit, but that assignments are often updated throughout the day, as some households take longer or shorter amounts of time than expected to interview. Performance is evaluated continuously throughout the survey round using an "interviewer progress sheet," shown in Figure 2.1. The supervisors keep one progress sheet per interviewer throughout the fieldwork process and update the sheet after each cluster to ensure that each member of their team is keeping up with the assigned workloads. I therefore examine number of completed interviews as a very highly-monitored measure of interviewer productivity.

The DHS has several other mechanisms in place for monitoring interviewer performance. As mentioned previously, field editors regularly observe individual interviews (they are responsible for one observation session per day). Supervisors conduct one reinterview per cluster to ensure that interviewers are not engaging in common practices meant to lighten workloads, such as interviewing smaller households that were not selected or classifying eligible individuals as ineligible for individual interviews by misreporting their ages. Finally, field editors check questionnaires in the field, providing an additional check on the quality of data produced by each interviewer. The data quality problems observed in the final data are flagged in secondary processing after fieldwork is complete; they are issues that were not caught during fieldwork. These flags for incomplete, inconsistent, or missing data comprise the outcome variables used to study production quality below.

2.3 Data

2.3.1 DHS Datasets

The dataset merges information on over 1.2 million individual DHS interviews with daily gridded global weather data from the Princeton Meteorological Forcing Dataset. I use data from all DHS survey rounds that took place before the end of 2010 that collected GPS information on survey clusters. This creates a dataset with 46 countries and 92 survey rounds.[14] Figure B1 shows the location of the countries in the sample. The dataset spans Africa, Asia, Latin America, and a few Eastern European countries. By and large, data collection takes place in countries that are warm on average (the average interview takes place on a 74 degree day), though the dataset includes a few cold, mountainous regions (in the Andes, Himalayas, etc.). Each DHS interview contains information on start time and end time in minutes, the interviewer's unique identifier, the supervisor and field editor's unique identifiers, a host of information on the respondent, and data quality information. As mentioned previously, the data quality indicators are added after fieldwork during the data processing stage, not by the field editor or supervisor. Many countries are surveyed more than once during the sample period of this study (that is, they have more than one DHS survey round in the dataset). The region of country fixed effects and interviewer identifiers are coded specific to certain survey round in the DHS.

[14]Table B6 gives detailed descriptions of each survey round included in the analysis.

The interviewers' unique identifiers and the interview start and end times are used to reconstruct the daily work schedules of each interviewer. The number of working hours in a day is calculated as the time between the start time of the first individual interview and the end time of the last individual interview conducted by the interviewer that day. This means that working hours exclude time spent commuting to and from the survey cluster, time spent locating the first household, and time spent conducting the first survey of household characteristics. It is possible that these activities are also affected by temperature, so estimated effects on working hours may be biased towards zero.

Table 3.1 contains summary statistics on the DHS data. The top panel contains individual-level information on female DHS respondents in the entire sample. The average age of female respondents is about 30, and over a third of respondents are illiterate. Under half have electricity and just over half live in houses constructed of formal materials (defined as a home with a formal wall, roof, and floor). About half of female respondents work, whether inside or outside of the home. An important aspect of this context is that very, very few households in developing countries in this time period would have had air conditioning. For example, only around 2 percent of households in India have air conditioning today, and the interviews in this sample took place between 1990 and 2010.

Key interview quality outcome variables are summarized in the bottom panel of Table 3.1. Summary statistics on counts of missing responses are shown in the first row. These are instances where the question should have been answered, but no response was present. I use this as a major outcome variable due to the possible implication of interviewer error (although this category also contains respondent refusals); these occur about once per interview on average, with a large standard deviation.

The DHS also creates data processing flags on certain key variables, mainly important dates. These variables flag instances where a major date, such as a birth date, is incomplete or inconsistent with another date given. Instances of body measurements that are flagged for falling outside of the expected distribution are also included, as are flags on duration variables, such as duration of breastfeeding, that are flagged for being inconsistent with another date given. The second data quality outcome variable is a count of these data quality flags in each interview: the second row of Table 3.1 summarizes this variable.[15]

The third data quality outcome variable is labelled "valid skips" and is defined as variables that are left blank due to the skip or flow pattern of the questionnaire. These are the "not applicable" or "not in universe" responses. The interview-level mean of this variable is very large, at over 6,300. This is due to the existence of many country-specific variables in the DHS, meaning that the majority of variables do not apply to any specific survey. All comparisons of data quality on days with different temperature are made within a survey round, meaning that survey-specific variation in which questions are asked is netted out. Figure B3 displays a section of the standard questionnaire for DHS Wave 7 that illustrates the difference between missing responses and valid skips. The section begins with instructions to check a previous question to verify whether the respondent is pregnant. If the respondent is pregnant, then the interviewer should move to question 312, and questions 303-311 will be marked as valid skips. If the respondent is not pregnant, then the interviewer should move on to question 303. If the respondent is marked not pregnant or unsure but question 303 is left blank, then question 303 is a missing (or invalid missing) response. Due to the skip patterns built into the questionnaire, interviewers may have an incentive over the answers

[15]More details on the construction of this variable can be found in Appendix B.1.

to questions that determine the total number of questions to be asked in the survey. For example, if they would prefer to ask fewer questions, they may erroneously write that the respondent is pregnant to avoid the contraception module. To the extent that this incentive is impacted by temperature, we would expect to observe an increase in the number of valid skips rather than invalid missing responses.

Finally, summary statistics on responses that were either inconsistent with another response or where the respondent did not know the answer to the question are shown in the fourth row of the outcome variable panel. These instances are coded separately from the invalid missing responses summarized in the first row, and they are coded mostly consistently in the DHS. However, there are some exceptions, as detailed in Appendix B.1. These exceptions are removed to the best of my ability. The average interview contains about 5 of these types of responses.

The final three rows contain interviewer-day level outcome variables. The average interviewer-day consists of 3 individual interviews over 4.5 hours. The interviewers complete about one interview per hour on average. In the main analysis, I also examine the the start times and end times of interviewers' days and the distribution of interview times throughout the day.

To validate the main outcome variables utilized in this study as measures of interviewer productivity, Figure 2.2 shows summary statistics on quantity and quality outcomes as they evolve throughout the survey round. The figure shows the results of regressions that examine the impacts of each additional day of interviewer experience, measured as the number of days between the day of interview and the first day of the survey wave, on the average number of interviews completed per hour by interviewers and missing responses per interview. The regressions include region of country fixed effects, and the magnitudes are interpretable as the effect relative to the first day in the round. The results show that there are significant returns to experience in both measures. For the number of interviews per hour measure, the returns are diminishing over time, as would be predicted by standard labor models. For number of missing responses, productivity increases over time at the beginning of the survey round, but then begins to deteriorate again towards the end of a survey round, consistent with DHS supervisor guidelines that caution team leaders that data quality often begins to slip at the end of a survey round, as the teams become tired and bored. Since the jobs are temporary, the continuation value of the job to the interviewer decreases towards the end of the survey round, which also potentially contributes to this result.

As a final note, interviewer productivity as measured by number of interviews completed per hour worked is correlated in the cross section with aggregate measures of labor productivity. Each additional $1,000 of GDP per person employed, as measured in the Penn World Tables, is associated with 0.012 additional interviews per hour on average, for example. The relationship is statistically significant.

2.3.2 Daily Weather Data

I merge information from each interview with the Princeton Meteorological Forcing Dataset, henceforth the Princeton Data. The Princeton Data is a reanalysis dataset, meaning that it combines a host of observational weather data from sources such as weather balloons, weather stations, and satellite images with a physics-based weather model. The model allows coverage to extend to areas with sparse observational data such as developing countries.[16] These data are available through 2010 every three hours for each 0.25 degree latitude/longitude increment. The dataset includes information on temperature, specific humidity, and pressure, which I use to calculate relative humidity and wet bulb temperature using standard meteorological formulas.[17]

Wet bulb temperature is a nonlinear function of dry bulb temperature (dry bulb temperature is the temperature reading usually displayed on daily weather forecasts) and relative humidity. A wet bulb reading corresponds to the temperature reading of a thermometer that has been wrapped in a wet cloth: the faster the moisture in the cloth evaporates, the lower the reading. To help the reader visualize the relationship between wet bulb temperature, dry bulb temperature, and relative humidity, Figure B2 displays the relationship between the three variables along with a few illustrative examples. The figure displays isometric dry bulb lines, with each line representing a fixed dry bulb temperature at varying relative humidities, represented by the horizontal axis. The combination of these two variables can be read on the vertical axis as a wet bulb temperature. Wet bulb temperature is always lower than dry bulb temperature, except at 100% relative humidity where the two readings are equal. The difference between wet bulb and dry bulb temperature is larger at higher temperatures.

The red points in the figure plot two examples of the differences between wet and dry bulb temperature. The August, 2017 average of daily average weather conditions are plotted for Las Vegas, NV and Houston, TX. The desert heat of Las Vegas manifests as a relatively high dry bulb temperature (over 90 degrees Fahrenheit), but, due to the low relative humidity, a comparatively low wet bulb temperature. The more muggy heat in

[16] Appendix B.1 gives more detail on the construction of this dataset.
[17] See Appendix B.1 for more details.

Houston creates a higher wet bulb temperature than Las Vegas despite the fact that the dry bulb reading is lower. The emphasis on relative humidity means that the ranking of temperatures in wet bulb and dry bulb temperatures are very different, with humid hot days in places like South Asia rated as more severe than the desert heat of places like the Arabian Peninsula or Sub-Saharan Africa by wet bulb temperature, but as less severe using dry bulb temperature. As I show in Section 2.5, wet bulb temperature is more predictive of worker productivity, which means that the effects of weather on productivity may be concentrated in more humid areas of the world.

I merge weather data to DHS interviews by locating the four grid points in the Princeton Data surrounding each DHS cluster (each cluster in this sample has 20 selected households on average and is represented in the DHS data by a single latitude and longitude) and then creating an average of the four grid points weighted by inverse distance between the cluster and each grid point. The weather variables are then specific to a day of interview/DHS cluster. For the main analysis, I collapse the eight daily readings into a daily average wet bulb temperature, but in Section 2.8 I show the results for alternative specifications of weather exposure. To more thoroughly control for local average climate, I also create 10-year averages of wet bulb temperature in a given month in a given cluster (2000-2010 averages).

Figure 2.3 displays summary statistics on the weather data merged with the DHS data. Panel A shows the average wet bulb temperature in each survey cluster sampled in the DHS surveys in the analysis, where the wet bulb temperature is divided into the same bins as in the main regression specification. The average is of daily average wet bulb temperatures during the interviews in the sample. The variation used in the analysis is within a region of the country, so identification comes from areas where there was more variation in wet bulb temperature across interviewing days within a geographical area in the main regression specifications.

Panel B presents this graphically; it shows the regional distribution of interviewer work days in each wet bulb bin used for the main analysis. As the figures make clear, a bit over half of all observations are in Sub-Saharan Africa (SSA). SSA's climate is often quite hot, but in most places it is a dry heat, which means that the wet bulb distribution in Africa is largely compressed to the middle bins. The days in the highest wet bulb bins are concentrated in humid locations of the world, such as South Asia and Latin America. Days in the highest wet bulb bin are quite rare, accounting for less than 0.1 percent of days in the sample, and these occur only in South Asia and Latin America in this sample. These days are projected to become more common with climate change, particularly in areas such as South Asia (Steven C. Sherwood and Matthew Huber, 2010).

2.4 Empirical Strategy

As is standard in the literature, I estimate the impact of temperature allowing for nonlinearities in the impacts of weather on productivity outcomes. More specifically, I estimate the effect of daily average wet bulb temperature falling into a certain bin on my outcome variable of interest, relative to an excluded bin of 50-60 degrees Fahrenheit.[18] I

[18] Recent work on temperature and economic outcomes has also begun to think more carefully about biases introduced by both short and long-term adaptation in panel estimates with fixed effects. These estimates, though they identify off of high-frequency weather variation, may be biased by the effects of expectations since weather is serially correlated and at least partially forecastable (Derek Lemoine, 2017). A significant body of recent work also uses higher order functional forms of temperature in regressions, allowing cross-sectional variation to re-enter the econometric models and therefore allowing the effects of temperature to vary with climate (see, for example Burke, Hsiang and Miguel (2015)). These regressions account for cross-sectional differences especially in cases where more of the identifying variation is across

control for a 10 year average wet bulb temperature in the survey cluster of the interview in the month of interview. Therefore, the thought experiment is to replace a 50-60 degree day with a day in the bin of interest, holding constant usual monthly place-specific temperature.

For my main productivity results, I run regressions at the level of the interviewer-day. The estimating equation takes the form:

$$y_{icprd} = \sum_{j} \beta_j \cdot Exposure_{cpd}(T_j) + \alpha ExpectedTemp_{cpm} + \theta_p + \rho Daylight_{cpd} + \\ \nu X_{rcpd} + \gamma DayinRound_{pd} + \epsilon_{icprd} \quad (2.1)$$

where y_{icprd} is the outcome variable (e.g. interviews per hour) of interviewer i in survey cluster c in country-region p interviewing a set of respondents r on day of interview d. β_j is the coefficient of interest and gives the effect of daily average wet bulb temperature falling in bin j on the day of interview on productivity, relative to the reference bin of 50-60 degrees Fahrenheit. I estimate the impact of 7 bins of wet bulb temperature: ¡30 degrees, 30-40 degrees, 40-50 degrees, 50-60 degrees, 60-70 degrees, 70-80 degrees, and ¿85 degrees.[19] As mentioned previously, I also control for a 10-year average of wet bulb temperature in the cluster of interview in that month. I control for region of country fixed effects: there are

space rather than over time (Pierre Merel and Matthew Gammans, 2017), although this means that omitted variables may also be a concern in these models (Marshall Burke and Kyle Emerick, 2016). Rather than allowing group means to enter identification through a higher-order term in the regression, this paper uses linear regressors and examines the potential for short and long-term adaptation by comparing the main results using detailed controls for place-specific seasonality, which isolate "surprising weather days" and those comparing productivity on hot and cold days in the cross section, which includes the effects of usual local climate.

[19] Temperature is defined specific to a survey cluster, and in certain cases, interviewing teams visit more than one survey cluster in a day (15 percent of interviewer-days). In these cases, I use the average temperature for the survey clusters, weighted by the number of interviews in each cluster.

nearly 800 regions in the sample, and regions are specific to a country and a survey round due to lack of consistency in coding of the region variable over time in the DHS.[20]

I also control for the number of daylight hours in the survey cluster on the day of interview.[21] Many of the main results examine hours worked and start time and end time of the day. Since daylight hours are correlated with temperature, this could be an important control, although in practice the results are nearly identical with or without it.

In addition, I include a series of respondent controls. Since the regression is run at the interviewer-day level, these are characteristics of the set of the respondents. I control for whether the respondent is illiterate, the number of own children under the age of five, household electrification, whether the house is constructed of formal materials, whether the respondent works, age, marital status, and whether the survey cluster is rural or urban. The main analysis is restricted to the sample of women: many control variables are only available for women, and the work of male interviewers is inherently different, since in many cases they don't conduct interviews in every household. Since it is possible that the set of respondents on hot days is endogenously chosen (if say, interviewers chose easier-to-interview households on hot days), there is a potential for these results to be contaminated by the "bad control" problem. That is, respondent characteristics may be outcome variables in themselves, so controlling for them may bias the results. However, as I show in Section 2.6, the results are very similar with or without respondent controls.

[20]In rare cases, interviewing teams cross region borders within a given day. In these cases, I assign them to the first region they visited, but the results look similar if I drop them altogether, as this occurs in less than one percent of interviewer-days.

[21]I use the sunrise/sunset calculations from the National Oceanic and Atmospheric Administration, available at: https://www.esrl.noaa.gov/gmd/grad/solcalc/calcdetails.html

The main specifications also include a control for the number of days elapsed between the start of the survey round and the date of the interview. Interviewing teams tend to begin the survey round in a central location, such as the country capital, and then spread throughout the country, which means that respondent characteristics are nonrandom with respect to date within the survey round. Furthermore, survey rounds start at nonrandom times of the calendar year: for example, the documentation for several survey rounds mentions planning the survey round to avoid a rainy season.[22] Finally, the standard errors are clustered at the region of country level.

For some of the main outcome variables in this study, such as counts of data quality problems and respondent characteristics, the regressions are run at the level of the individual interview. In this case, the estimation equation is similar to the interviewer-day regression. However, the respondent controls are for individual characteristics, rather than for the set of respondents.

Figure 2.4 maps the survey round in Nepal in 2006 in order to help the reader better visualize the regression strategy. In the figure, the interior boundaries represent the five regions for which the regression has fixed effects, while the different colors/shapes of the points on the map represent each of the 12 interviewing teams (defined by team supervisors) in the survey round. It is clear from the figure that each team is seen in multiple regions, with each individual survey cluster covered by a single team. In the dataset, the teams spend 2.7 days in an individual survey cluster, on average. There are on average 35 individual

[22] In practice, as I show in Table B2, hot days tend to occur later in the survey round on average, when interviewing teams are more experienced. Including the control for day of round magnifies my estimated effects, as I show in Section 2.8, implying that any remaining bias may work against finding any effect of temperature on productivity.

interviews in a survey cluster (36 in rural areas, 33 in urban areas), coming from 19.5 households.

2.4.1 Identification Assumption

This analysis estimates the impacts of temperature fluctuations on data production conditional on average local climate and other region-specific factors. The key identification assumption is that factors associated with data production are not correlated with within-region fluctuations in weather conditional on climate. Potential omitted variables could occur either at the level of the respondent or survey cluster or at the level of the interviewer or survey team. At the level of the respondent, the assumption is that the set of respondent households and associated characteristics (such as distance between households) is comparable on hot, mild, and cold days within a region of a country and conditional on climate. This assumption would not be violated, therefore, by respondents being less available during certain seasons due to the usual seasonality of agricultural activities, for example, but it would be violated by respondents being less likely to be home on an unseasonably hot day.

At the level of the interviewer/team, the assumption is that the composition of interviewer characteristics is comparable within a region on hot, mild, and cold days conditional on climate. This would not be violated by survey progression being scheduled to avoid hot months of the year, but it would be violated by certain types of interviewers being less likely to show up to work on surprisingly hot days.

I investigate the identification assumptions empirically in Section 2.6 and Section 2.8, showing robustness to using interviewer fixed effects and controlling or not controlling for interviewer experience as well as demonstrating that selection on observable characteristics of respondents does not appear to drive the results.

2.5 Results

2.5.1 Effects of Temperature on Productivity

Results from Equation 2.1 are shown in Table 2.2 and Figure 2.5. Table 2.2 displays the results using several specifications for contemporaneous weather. The outcome variable throughout the table is the number of interviews completed per hour worked, where an observation is an interviewer-day. Column 1 examines the impacts of average dry bulb temperature in the day of interview, where each coefficient demonstrates the impact of replacing a day with an average temperature between 50 and 60 degrees with a day in the bin of interest. The results suggest a negative impact of high heat, but they are not statistically significant.[23] Columns 2 and 3 display the results from a single regression investigating the impact of dry bulb temperature interacted with an indicator for high humidity, defined as humidity on the day of interview falling above the 75th percentile of the 10-year average of humidity in the survey cluster in the month of interview. These regressions also include estimates for the effects of dry bulb temperature on low-humidity days as well as for the main effect of humidity. The results are highly significant for high humidity days but not low humidity days, suggesting that humidity plays an important role in the effect of temperature on productivity per hour. Finally, column 4 examines the impact of wet bulb temperature. The table also displays F statistics jointly testing the statistical significance of the dry bulb temperature bins, dry bulb X humidity bins, and wet bulb bins, respectively. The F statistics are highest in the specification with wet bulb temperature, despite the fact that there are

[23]However, the results are statistically significant and larger if climate controls are not included.

fewer coefficients, suggesting that wet bulb temperature has the most predictive power for this measure of productivity. Therefore, I examine the impacts of wet bulb temperature throughout the rest of the paper.

Figure 2.5 shows the effects of wet bulb temperature on number of interviews completed per hour in figure form, corresponding to column 4 in Table 2.2. Each point in the figure gives the impact of replacing a day with an average wet bulb temperature between 50 and 60 degrees Fahrenheit with a day in the degree bin specified on the horizontal axis. In this figure, days colder than 50-60 degrees seem to improve productivity per hour worked, especially in the 30-40 and 40-50 degree bins. In contrast, high temperatures, in particular the ¿85 degree bin, are detrimental to productivity, with interviewers experiencing a 20% loss in interviews per hour on these exceptionally hot and humid days. Figure 2.5 provides strong evidence of harmful effects of high temperatures on data production as well as evidence of beneficial effects of cooler temperatures. The effect appears to be quite linear in temperature until the highest bin, where there is a trend break as productivity drops off sharply.

2.5.2 Temperature, Data Quality, and Interviewer Incentives

Given the strong evidence that temperature affects data production, I now turn to interviewer incentives to examine whether interviewers respond to temperature in a predictable manner. As mentioned in Section 2.2, DHS supervisor and field editor guidelines suggest that the quantity of data produced, especially in terms of number of interviews completed, is more easily observable to supervision than quality. Given that temperature may increase the disutility of effort or decrease the marginal benefit of effort in terms of productivity, we may expect temperature to produce a differential change in productivity for quality vs. quantity of data production.[24]

I first examine the impact of temperature on the measure of productivity most observable to supervisors: the number of completed interviews. Figure 2.6 gives the results from Equation 2.1 where the outcome variable is number of interviews completed in a day. Interviewers do not seem to perform worse on this measure on very hot days; the coefficient on 85 degree wet bulb days is actually positive, though statistically insignificant. Given that interviewers are not paid by the hour, this suggests that from the employer's perspective, productivity (at least, in terms of quantity of data produced) per dollar of wages does not decline on hot days. However, productivity per hour of the interviewer's time declines, as shown in Figure 2.5, suggesting that interviewer welfare may be negatively affected by high heat through a loss of leisure hours. This is explored further in Section 2.5.3.

To further probe into whether productivity declines on less observable tasks, I use a framework similar to Equation 2.1 to examine whether measures of data quality respond to high temperatures. These regressions are run at the level of the individual interview. Figure 2.7 shows the results of regressions using counts of data quality flags and missing responses, respectively, as outcome variables. For both measures, the coefficients on the hottest wet bulb bins are positive and statistically significant, indicating these types of mistakes are more common on hot days. These two pieces of evidence together suggest that performance suffers more on hot days on measures that are less easily scrutinized by the supervisor.

[24]In Appendix B.2, I show a simple model that produces this prediction in the case where interviewers' probability of job retention depends more on quantity of production than quality and temperature increases the disutility of effort.

Table B3 shows the results for the missing responses and quality flags variables shown in Figure 2.7, but also for counts of valid skips and counts of inconsistent or "I don't know" responses. The results on valid skips, displayed in Column 2, suggest a positive effect of both very cold and very hot temperatures on number of blank questions, although the impacts of hot temperatures are not statistically significant. The counts of don't know and inconsistent responses do not show a clear pattern according to temperature on the day of interview, suggesting that respondents' ability to respond correctly may not be significantly affected by wet bulb temperature.[25]

Another feature of the interviewers' incentive scheme that may affect the results is the fact that their main incentive to perform well, due to the fixed wage contract, comes from any risk of losing the job and the continuation value associated with it.[26] Interviewer jobs are temporary, so the continuation value of the job (i.e. the quantity of wages that would be lost if the interviewer were fired) should decrease as the survey round progresses, which could amplify any negative effects of heat on productivity. On the other hand, experience may be protective or may lead workers to learn to take more adaptive actions. Figure 2.8 examines how experience interacts with temperature by interacting each wet bulb bin with a measure of how many days the interviewer has worked on the survey round. The outcome variable is number of interviews completed per hour worked. In fact, the effect of hot weather significantly increases with experience.[27] As shown in Figure B6, the impact of experience

[25]However, examining the impact of dry bulb temperature on this outcome variable yields a positive and significant effect. This could be driven either by interviewers or respondents, since interviewers may have an incentive to write that the respondent said "I don't know" in response to a question in order to lighten their workloads.

[26] The model in Appendix B.2 additionally predicts that an increase in the continuation value of the job should be associated with a reduction in the negative effect of temperature on interviewer effort.

[27]There may be selection into levels of experience, if, for example, less experienced workers are more likely

at high temperatures is driven by working hours: more experienced interviewers are in the field for more hours on hot days. This could be consistent with a learning effect: more experienced interviewing teams learn to cope with the slowing pace of worker activity by working longer hours. Or, it could be that workers with lower continuation value of the job differentially allow themselves to slow down on hot days due to weaker incentives to perform well. On the other hand, the impact of experience on reactions to cold temperature are driven by the numerator: experienced interviewers conduct fewer interviews on cold days.

2.5.3 Temperature, Labor Supply, and Time Allocation

The fact that number of interviews completed per hour declines on hot days with little change in the numerator suggests a lengthening in working hours. As mentioned previously, the opposite phenomenon has been observed in the U.S., where workers appear to work fewer hours on hot days (Graff Zivin and Neidell, 2014). Given these differences in results, in this section, I examine the types of adjustments workers are making to their workdays on hot days in developing countries. I first examine summary statistics on the starting times and ending times of interviewer days in my sample by wet bulb bin. The results, displayed in Panel A of Figure 2.9, show that interviewers' days begin monotonically earlier as the wet bulb temperature rises, while end times appear to be less responsive. The result is that interviewers work more hours on hot days on average, and the difference between the hottest and the coldest bin is quite stark: work days are about four hours shorter on days with wet bulb temperatures less than 30 degrees than for days over 85 degrees wet bulb. This is

to be workers who were brought in later in the survey round to replace those who were fired. Figure B5 examines the same relationship but using interviewer fixed effects. The results are very similar, suggesting that selection into experience does not drive these results.

consistent with anecdotal evidence from developing countries suggesting that workers may have to work longer hours in order to hit daily production targets (Kjellstrom et al., 2016*b*).

To more closely investigate this pattern, Panel B of Figure 2.9 gives kernel density plots of individual interview start times throughout the day for each wet bulb bin. The figure shows the same pattern as the previous one, where interviewers start their morning interviews earlier but don't end their afternoons significantly earlier. From this figure, it also becomes clear that the interviewers take longer mid-day breaks on hotter days. Therefore, it appears that interviews are allocated towards times of day with more pleasant temperatures: in the morning and later afternoon on hot days and mid-day on cold days.

These figures draw from summary statistics, so there are many possible explanations for the patterns shown in the average interviewer schedules. One such explanation could be the relationship between daylight hours and temperature: if the sun rises earlier on warmer days on average, then this could account for the earlier start times, rather than temperature itself. However, as Table B5 shows, this does not appear to be the case. The table shows correlations from regressions without controls or fixed effects; daylight hours have the expected negative correlation with start time: interviewing days start earlier on days with more daylight. However, when added as a control, daylight hours do not appear to mediate the relationship between temperature and start time.

To separate the causal effect of temperature on working hours from other place-specific factors, in Figure 2.10, I investigate the relationship between wet bulb temperature and start time using multiple specifications. The dark blue line gives the cross sectional relationship between temperature and start time; this is the same relationship displayed in Figure 2.9, except in regression form with clustered standard errors and non climate-related controls added. The red line adds controls for usual wet bulb temperature in the month of interview for that cluster (this is the 10-year average used in the main specification). The slope is significantly shallower in this specification, implying that much of the relationship between temperature and start time is due to a relationship between climate and start time; that is, places that are hotter on average start work days earlier on average (at least in times of year that are hot). The light blue line gives the results of the full specification, with survey round by region of country fixed effects and other controls. The coefficient on the hottest wet bulb bin remains negative and statistically significant in this specification, but the relationship across the rest of the distribution has disappeared. Again, these results suggest that most of the response along this margin is to usual seasonality, rather than "surprise" weather days. This suggests that a lengthening in the work day may be a form of long-term adaptation in these settings: workers accommodate the slowing pace of production by starting earlier. An implication of this is that in the longer run leisure hours and productivity may be more negatively affected by high temperatures than in the shorter run. In addition, workers may start earlier on days they *expect* to be hot; this ability to forecast may be factored out in specifications that narrow in on surprising weather days.

Finally, in Figure 2.11, I examine the contribution of increases in interviewing duration to the increase in overall time in the field. I find that even in the main regression specification, with region of country fixed effects and climate controls, time spent throughout the day conducting interviews increases significantly on hot days relative to mild days. Therefore, interviewers are both spending more time in the field on hot days and more time actively conducting interviews.

The previous results show that workers increase labor supply along the intensive margin on hot days, but another way they could respond is through adjustments along the extensive margin. That is, they could not conduct interviews at all on days with extreme temperatures. In Figure 2.12 I investigate this possibility by running a version of Equation 2.1 using a dummy variable for whether an interviewer was observed conducting interviews on a given day. I include in the sample all days between the interviewer's first and last interviewing date in the survey round.[28] The results of this exercise show that interviewers are significantly less likely to work on hot days relative to mild days and more likely to work on colder days. The magnitudes are large, hovering around 20 percentage points at the highest temperatures. There are several potential channels for these effects. Interviewing teams may schedule travel days (between survey clusters) to coincide with unpleasantly warm temperature days in order to avoid the effects on interviewing performance. Interviewing teams also generally take regular breaks so that the interviewing teams can return home and visit their families; it may be that these are also strategically scheduled. Or, it may simply be that individual interviewers are less likely to show up for work on hot days. In any case, it appears that adjustments in labor supply on the extensive margin are one manner in which interviewing teams avoid some productivity and utility consequences of extreme temperature.

[28]This exercise is complicated by the fact that interviewers are observed in many different places throughout the survey round. I assign each non-working day the weather from the most recent survey cluster visited by the interviewer. To minimize the possibility that they have travelled large distances since their last observed interview, I limit the sample to interviewers observed working within the last 2 days, though the results look similar if that period is extended.

2.6 The Role of the Respondent

In this section, I discuss the potential impact of respondent behavior on the results. Household survey interviews are a unique type of economic production: the respondent plays a key role in the production outcome, but they are in a sense doing the interviewer a favor by participating. That is, they are not participating in the interviews in order to garner economic gains for themselves. This means that, from a generalizability standpoint, the interviewer's response to temperature is most interesting. With that in mind, I discuss the specific mechanisms through which respondent behavior could influence the results in this section and investigate those mechanisms to the degree possible using the data.

The respondent could influence the results in three ways. First, if fewer respondents are readily available for interviews, the interviewers may spend more time walking between respondent households looking for respondents, resulting in fewer completed interviews per time worked. In about 10 percent of cases, interviews in the DHS take more than one attempt to complete, and while DHS overall response rates are very high (usually over 90 percent) not all interviews are completed. As shown previously in Figure 2.11, the increase in working hours is mainly driven by lengthening interviews, suggesting that the results are not driven by an increase in the time spent looking for respondents.[29]

Secondly, an effect on nonresponse could manifest itself as a selection effect. If certain types of respondents are less likely to be available on hot days, then there could be selection into the sample of respondents on hot days. This could drive the results if the pool

[29]Unfortunately, my ability to observe nonresponse directly in the data is limited. As mentioned previously, the data has a variable that specifies how many visits were necessary to complete the interview, but I am unable to observe when the previous attempts took place; it is possible that the unsuccessful attempts are on a different day from the successful one. Therefore, I don't investigate nonresponse as an outcome directly.

of respondents observed on less hot days is comprised of people that are easier to interview efficiently than the pool of respondents observed on hot days. While I'm unable to fully observe which respondents are "easy" to interview, I can investigate whether the sample of respondents on hot days looks different from the sample on mild days based on observable characteristics. I conduct this exercise with a number of observable characteristics in Table 2.3. This table contains the results of regressions, run without respondent controls, using individual respondent characteristics as the outcome variable of interest.[30] Few characteristics show a statistically significant response to temperature. There is some evidence that on hot days, respondents who work may be more likely to be interviewed, as are respondents with more children under the age of five, but the signs are inconsistent across bins. Respondents on very hot and cold days appear to be older, on average. On cold days, I also see respondents with more children, as well as fewer respondents with a formally constructed house. For all of the variables except for sex, the regressions are run only on the sample of female respondents. Since interviewers specialize, in general, in either female or male respondents, using respondent sex as the outcome variable is more of a check on the relative productivity of female and male interviewers on days with different temperature. The evidence suggests that more male respondents are seen on cold days.

Given the presence of some selection into the sample based on observable characteristics, one way to test the impact on the results is to compare the results with and without controls for the same observable characteristics. If selection into the sample based on these observable characteristics were impacting the results, then we would expect the results to

[30] The asset index is a count of ownership of several household assets. These are: a radio, a car, a motorcycle, a bicycle, electricity, a television, a refrigerator, and piped water.

look different if the controls were removed. However, as I show in Figure B4, removing these controls does not affect the results. While it is impossible to investigate empirically whether selection on unobservable characteristics occurs or impacts the main results, this evidence is consistent with the results not being driven by selection on respondent characteristics.

The final way that respondents could have an impact on the results is more difficult to investigate empirically. Respondents, being also exposed to temperature conditions during interviews, may simply behave differently during the interview. They may be less able to remember important details of their lives due to the cognitive impacts of high heat, for example. Or, they may be less cooperative with the interviewer. Either of these effects would mean that obtaining correct answers in a timely manner would become more difficult on a hot day, even without any direct effect on the productivity of the interviewer. Because every respondent is only interviewed once in this dataset, it is impossible to disentangle the relative impacts of temperature on interviewer and respondent behavior.

One possible way to look for suggestive evidence on the effect of temperature on respondent behavior is to look for interaction effects between respondent characteristics and temperature. If the effects of temperature varied with respondent characteristics, it may indicate that respondents do have significant impacts on the results through their behavior (and that this varies by type of respondent), although it's also possible that interviewers' productivity varies with type of respondent. I investigate interactions between temperature and respondent characteristics in Figure B7, using the main outcome variable of number of interviews completed per hour. There is little evidence of heterogeneity, although a few interaction terms are statistically significant. Overall, however, the effect on number of interviews per hour persists regardless of the makeup of the set of respondents.

2.7 Heterogeneity in Main Results

The main results of this paper examine the impact of temperature on the productivity of survey interviewers across 46 developing countries over two decades. One major advantage of using this dataset is the comparability in the productivity measures used in this analysis across a huge variety of geographic contexts. This section investigates whether the response of DHS interviewing teams to weather conditions varies by area of the world or institutional context.

Supervisors and field editors in the DHS have specific instructions on how to monitor and evaluate interviewer performance, so these procedures are quite constant across the contexts of the countries in this analysis. However, the surveys are implemented in a huge variety of institutional settings. Factors such as generosity of pay, overall quality of survey oversight, and the desirability of interviewer jobs, to name a few, likely vary by survey round. These factors may provide additional variation in the strength of interviewer incentives, but they are difficult to observe, given lack of DHS records on pay levels among other things. One candidate proxy for strength of interviewer incentives is the quality of public sector management in the implementing country. In Figure 2.13 I examine interactions between daily average wet bulb temperature and the quality of public management as measured by the World Bank Country Policy and Institutional Assessment (CPIA).[31]

While these results should be viewed as suggestive, there are significant interaction effects.[32] Figure 2.13 plots interaction effects for number of interviews completed in a day

[31] More information on this index is available at: https://data.worldbank.org/indicator/IQ.CPA.PADM.XQ?view=chart. The index is available from 2005-2010 for 28 out of the 46 countries in the main sample.

[32] Interaction effects with other indicators of institutional quality, such as the Transparency International

(Panel A) and minutes spent working (Panel B), respectively. Interviewers complete fewer interviews on hot days in countries with lower quality public sector management, in contrast to countries with high quality public sector management, where there is no discernible impact on this measure of productivity. This is consistent with the previous evidence suggesting that the strength of incentives for this particular measure of productivity is important in determining the impacts of temperature. Interviewers in countries with high quality public sector management work for more time on hot days in order to complete the same number of interviews, whereas interviewers in settings with low quality management work fewer minutes on hot days, consistent with the impact on quantity of data production.

Table B4 examines the effects of wet bulb temperature on number of interviews completed per hour for each of 6 world regions; each column of the table represents a separate regression.[33] Interestingly, in regions where maximum wet bulb temperatures are lower, the negative effects of temperature appear lower in the distribution. It appears that the negative impacts of wet bulb temperature in the 70-80 and 80-85 degree bins are driven by regions such as Africa, the Middle East and North Africa, and East Asia Pacific, where wet bulb temperatures never exceed 85 degrees in this sample. This may suggest long-term adaptation to place-specific normal weather conditions, where productivity falls at the upper end of the temperature distribution, at whatever temperature that may be. Finally, the overall results from Figure 2.5 suggest a strikingly linear relationship between productivity and wet bulb temperature until 85 degrees wet bulb, when efficiency falls sharply. Figure B4 suggests

Corruption Index, have yielded noisier results. Additionally, there are well-documented correlations between institutional quality and other factors that may significantly affect the results, such as long-term climate (Dell, Jones and Olken, 2014).

[33] These regions are Africa, East Asia Pacific, Europe and Central Asia, Latin America and the Caribbean, the Middle East and North Africa, and South Asia.

that that shape may be place specific. The positive effects of cold temperatures appear to be driven by Europe and Central Asia, while cold days in regions like the Middle East and North Africa and South Asia appear to have negative productivity effects, leading to more of an inverse-U-shaped relationship between temperature and productivity.

The final three columns of Table B4 split the sample into thirds based on annual average wet bulb temperature to examine heterogeneity by usual climate. The temperature distributions of the three groups do not fully overlap, but there is some indication that the effects for 70-80 degree days are largest in the coolest third of clusters. This is consistent with evidence from Patrick Behrer and Jisung Park (2017), who find that the impacts of heat on labor income are largest for cooler climates. Cold days do not occur in the warmest third of clusters, so examining the effect of cold days in warm clusters is not possible in this sample.

2.8 Robustness and Alternative Specifications

I next examine the robustness of my main results, on the number of interviews completed per hour, to the choice of specification. In Section 2.6, I showed that the main results are robust to using or not using controls for respondent characteristics. In Figure 2.14, I show the robustness of the main results to other variations on the control variables. First, as discussed in Section 2.4, there is a correlation in the sample between the occurrence of hot days, even after controlling for expected temperature, and how far into the survey round the interviews occur, since survey rounds tend to begin during milder seasons. Figure 2.14 shows that this relationship biases the results towards zero: when I remove the control for the day of the round, the results become smaller, as shown by the dark blue line. The red line in

Figure 2.14 shows the results of a specification that replaces region of country fixed effects with interviewer fixed effects. If it is the case that more productive interviewers are more likely to be absent on hot days or that more productive teams are less likely to go to the field on hot days, selection on the sample of interviewers may drive the results. The results become smaller and marginally lose statistical significance for a few of the bins (the standard errors also grow substantially); however, the result suggests that overall, it is unlikely that the results can be explained by factors such as selection into the interviewer sample on hot days.

The weather data used for this analysis are available every 3 hours, or 8 times daily; in the main analysis I reduce the 8-daily observations to a daily average wet or dry bulb temperature as the independent variable of interest. Whether daily average temperature is the most relevant measure of weather for the outcome variables I study is an empirical question. Therefore, in Figure B8, I examine the main results using maximum wet bulb temperature, defined as the highest of the eight daily readings. While the coefficient on the hottest maximum wet bulb bin is negative, it is not statistically significant. The difference in responses to the two measures is driven by a lack of response in working hours to maximum wet bulb temperature, suggesting that scheduling is more responsive to the overall temperature conditions of the full day than to the maximum wet bulb temperature. The results look more similar to the results derived using daily average wet bulb temperature when 10-year average climate controls are not included, so another possible explanation for the disparity is that surprisingly hot maximum wet bulb temperatures are more difficult to forecast in these contexts.[34]

[34]An alternative could be to assign temperature based on the times when interviews took place. However,

2.9 Discussion

The main results of this paper find that DHS data production is negatively impacted on very hot and humid days relative to mild days. The number of interviews completed per hour worked declines, driven by an increase in working hours with no corresponding increase in output, and data quality issues increase in frequency. This section first puts this evidence in the context of existing literature and then discusses the generalizability and implications of the new evidence presented here.

The main results of this paper align with previous anecdotal, laboratory, and cross-sectional work finding that workers become slower to complete tasks on hot days. Previous work on panel data in a factory setting found that productivity dropped off at wet bulb temperatures higher than 66 degrees (Adhvaryu, Kala and Nyshadham, 2018). This paper suggests a more muted response once local climate is taken into account until 85 degrees wet bulb, after which production falls precipitously.

The mechanism behind this effect is a longer schedule to hit the same production target, consistent with anecdotal evidence that workers in developing countries work longer days when it is hot, sometimes with no change in pay (Kjellstrom et al., 2016*b*). It contrasts, however, with evidence from the U.S. that suggests that outdoor labor supply decreases in response to high temperatures. The contrast may be due to worker incentives in this setting: workers work longer hours in order to finish the same number of interviews, potentially because that is how their performance is evaluated by supervisors. Furthermore, there is

the interviewing schedules are endogenous and are in fact the major margin of adjustment for number of interviews completed per hour worked. With this in mind, I focus on weather specifications that do not take into account when in the day the interviews took place.

heterogeneity in that response in this setting: interviewers complete fewer interviews on hot days in countries with lower quality public sector management, consistent with the idea that the strength of incentives is important in determining the supply of effort.

The lengthening in schedule in the DHS dataset is more pronounced when local climate is not controlled for, suggesting that working longer days may be a form of adaptation in developing countries, as workers may avoid the hottest hours at the expense of spending more time on the job. This is a form of adaptation that may come at significant cost, both in terms of worker utility and in terms of productivity. Workers may be maximizing utility by allocating work time towards more pleasant hours, but there is presumably some utility cost of staying in the field for more hours. The duration of each individual interview also increases on hot days. Furthermore, this lengthening of the work schedule drives a decrease in production per hour worked.

To contextualize the size of these effects, Figure 2.15 visualizes the number of additional interviewer hours in the field that were required to complete the most recent DHS survey round in each country, relative to a counterfactual scenario in which all interviewing days had an average wet bulb temperature of 50-60 degrees. I calculate the actual number of interviews completed per hour worked by each interviewer on each day of the survey round and then calculate what the number of interviews completed per hour would have been under the counterfactual scenario. I then calculate the implied difference in working time and add this measure up for each interviewer-day in the survey round. The calculation thus assumes that no interviews are re-shuffled between days in this counterfactual scenario. The results suggest that the average DHS survey required almost 5 percent more interviewer hours than under the counterfactual scenario of mild weather throughout the round.

The results of this paper also suggest that production quality may be another margin of adjustment on hot days. Quality is, in this setting as in many others, more difficult to observe than quantity of output. Temperature may act as a source of variation in disutility of effort, causing effort to decline on hot days, especially on tasks that are less easily observed by the supervisor. To my knowledge, this is the first study to find an impact on production quality, and the results suggests that contracts and incentives in workplaces with incomplete monitoring may be less effective as climates warm.[35]

This evidence is derived from an unusual setting: data production is not a major sector of any economy.[36] Furthermore, workers with a secondary education who also work outdoors are a relatively small group in developing countries. However, workplace settings that involve daily production targets, exposure to non-climate controlled settings, and incomplete monitoring are commonplace. In addition, this setting, due to the standardization in interview questions and procedures across contexts and over time makes it possible to examine the impacts of temperature on labor productivity with a geographical breadth difficult to reach with other data sources.

2.10 Conclusion

In this paper, I provide estimates of the impacts of temperature on individual-level worker productivity and effort allocation, examining the effects of wet bulb temperature in

[35]However, James Archsmith, Anthony Heyes and Soodeh Saberian (2018) find a significant impact of air pollution on the quantity of umpire errors in baseball.

[36]However, workers with experience as interviewers are not uncommon in some developing countries. According to data collected by (Jobiba Chinkhumba, Susan Godlonton and Rebecca Thornton, 2014), one in ten men overall and one in four men with a secondary education ages 18-40 in Lilongwe, the capital of Malawi, had ever worked as an interviewer.

46 developing countries. I exploit a unique setting in which worker productivity data are available for a job that is standardized across a diverse set of contexts: DHS interviewers. I find that data production is significantly impacted by hot and humid weather: the number of interviews completed per hour worked declines by 20 percent of the mean on the hottest days, driven by an increase in working hours. In addition, quality of the data deteriorates on hot days, suggesting that workers, under the strain of high temperatures, allocate their effort to tasks that are more observable by their supervisors. While data production is a unique context, these results suggest that certain types of adaptation to heat and to disutility of effort more generally, may magnify the consequences for productivity. In addition, the results build on a recent literature showing the importance of humidity in the impacts of temperature, suggesting that hot and humid areas of the world may bear the largest productivity impacts of climate change.

Figure 2.1: Sheet for Evaluating Interviewer Performance

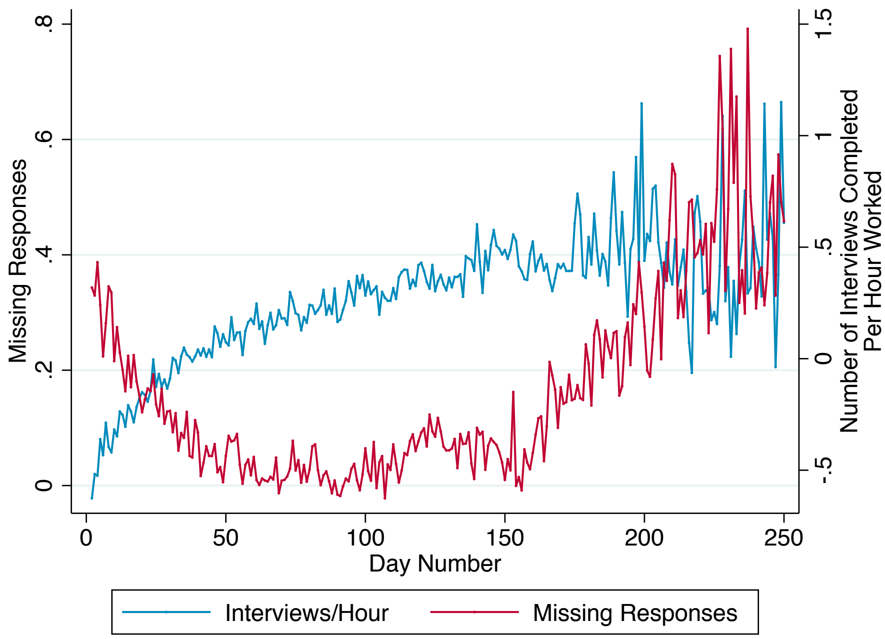
DHS INTERVIEWER PROGRESS SHEET
(USE A SEPARATE SHEET FOR EACH INTERVIEWER)

INTERVIEWER NAME ____________________________

	HOUSEHOLD QUESTIONNAIRES				WOMAN'S QUESTIONNAIRES				MAN'S QUESTIONNAIRES			
	NUMBER COMPLETED		NUMBER NOT COMPLETED		NUMBER COMPLETED		NUMBER NOT COMPLETED		NUMBER COMPLETED		NUMBER NOT COMPLETED	
CLUSTER NUMBER (1)	PER CLUSTER (2)	CUMULATIVE (3)	PER CLUSTER (4)	CUMULATIVE (5)	PER CLUSTER (6)	CUMULATIVE (7)	PER CLUSTER (8)	CUMULATIVE (9)	PER CLUSTER (10)	CUMULATIVE (11)	PER CLUSTER (12)	CUMULATIVE (13)

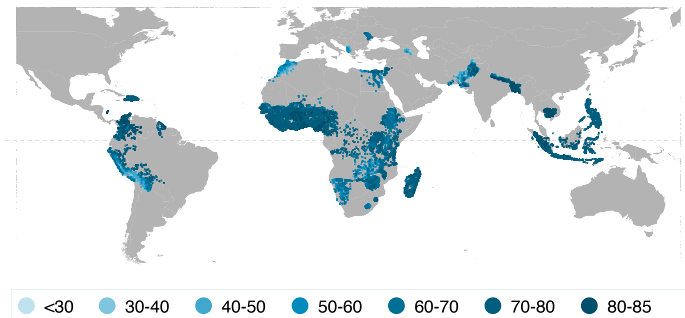
Note: This figure contains the sheet used in the DHS by supervisors to evaluate interviewer performance throughout the survey round. The sheet is intended to help the supervisor ensure that each member of his or her team is keeping up in terms of interviewing workload.
Source: DHS Program Supervisor's and Editor's Manual, July 2015.

Figure 2.2: Outcome Variable Validation: Productivity Measures Show Returns to Experience

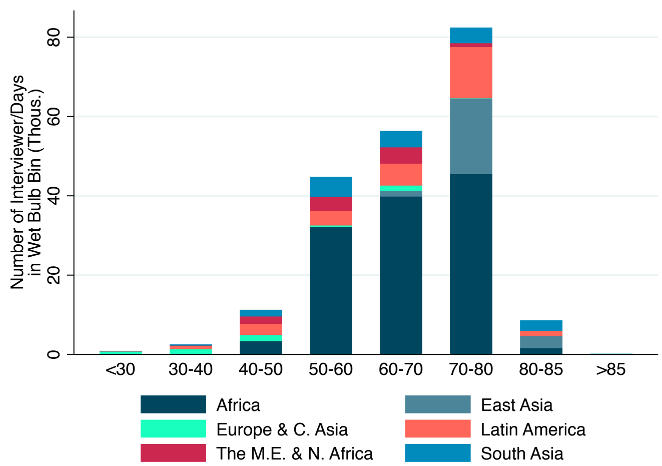


Note: This figure shows summary statistics on the average of number of interviews completed per hour and counts of missing responses as the survey rounds progress and interviewers gain experience, relative to the first day of the survey round. The horizontal axis is the number of days completed in the survey round (days between the day of interview and the first day in the survey wave). The figure shows the results of regressions that additionally control for region of country by survey round fixed effects.

Figure 2.3: Average Wet Bulb Temperature of Interviews in Each Cluster



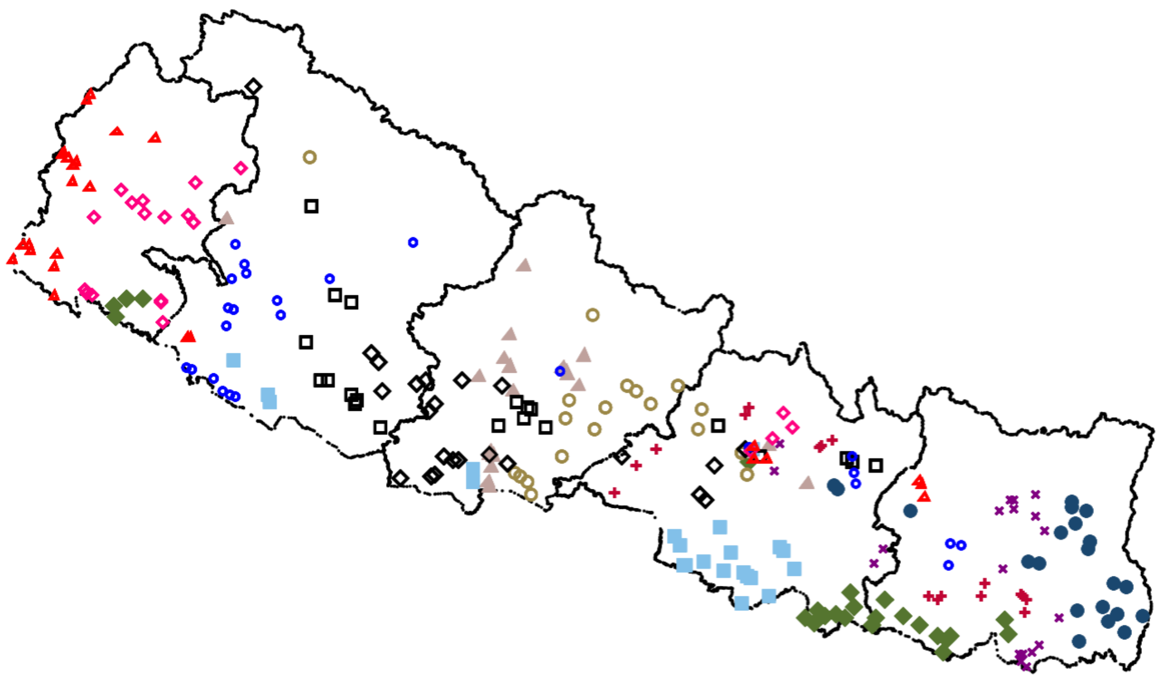
(a) Average Wet Bulb Temperature of Interviews in Each Cluster



(b) Distribution of Interviewer Days by Region and Wet Bulb Bin

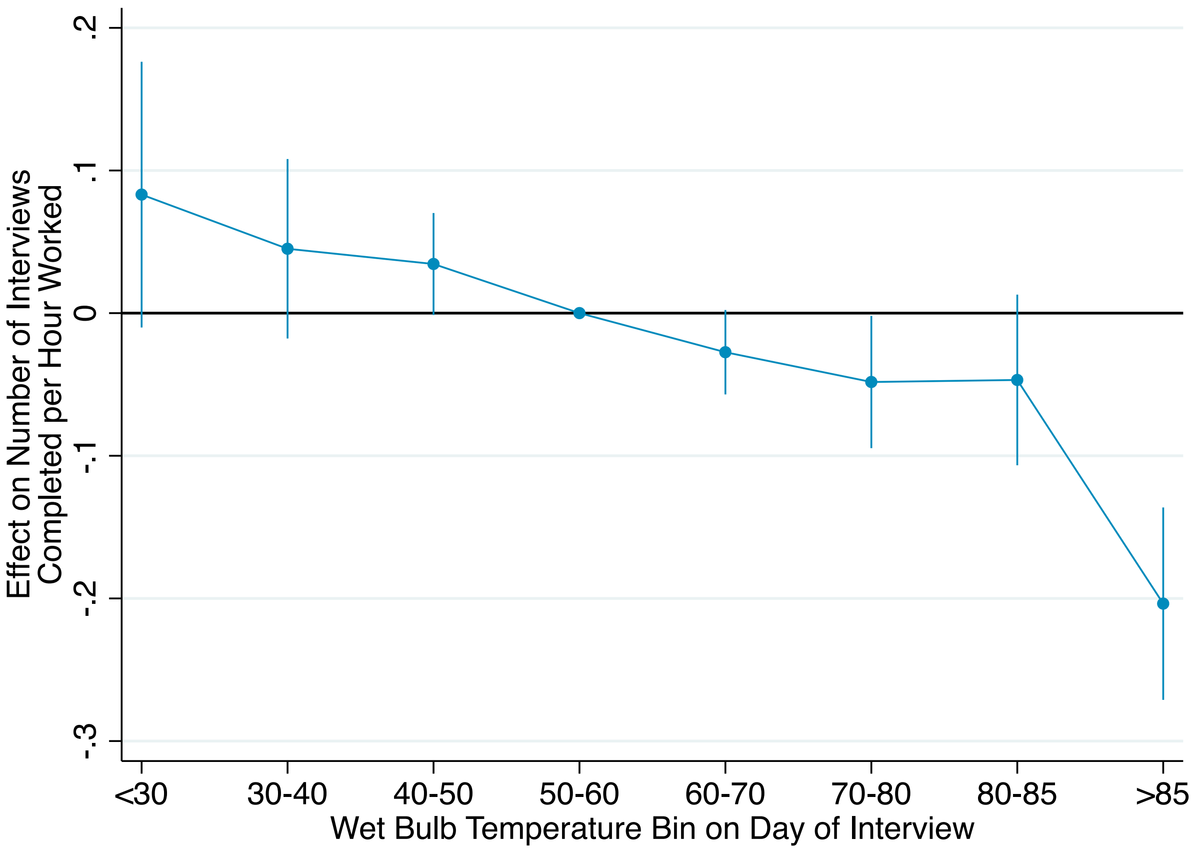
Note: Panel A shows the average of daily average wet bulb temperatures experienced in interviews in each cluster in the analysis. The shades of blue denote the bin this average falls into, with lighter shades representing cooler temperatures and darker shades warmer temperatures. In a few cases (3 percent), clusters with the same longitude and latitude were interviewed in multiple survey waves. In these cases, averages from the most recent round are given. Panel B shows the regional distribution of interviewer days in each wet bulb bin used in the main analysis. Each bar shows the number of interviewer/days in each world region taking place in a specific wet bulb bin, in thousands.

Figure 2.4: Summary Statistics: Visualization of the 2006 Nepal Survey Round



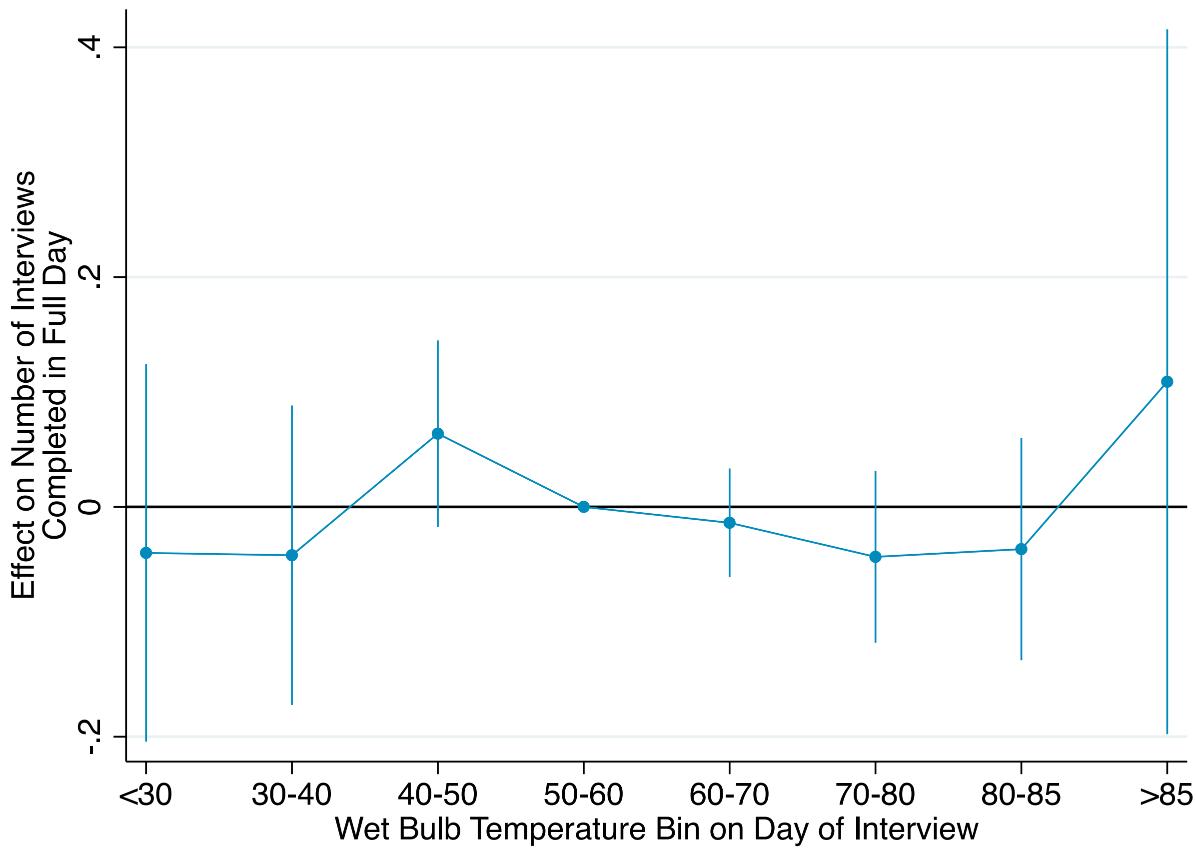
Note: This map gives a visual depiction of the 2006 Nepal survey round. The 5 regions depicted in the map are the regions used as place fixed effects in the analysis. Each color/shape indicates a different interviewing team, and the points on the map show where in Nepal the interviews took place.

Figure 2.5: Main Result: Number of Interviews Completed Per Hour Worked Declines on Hot Days



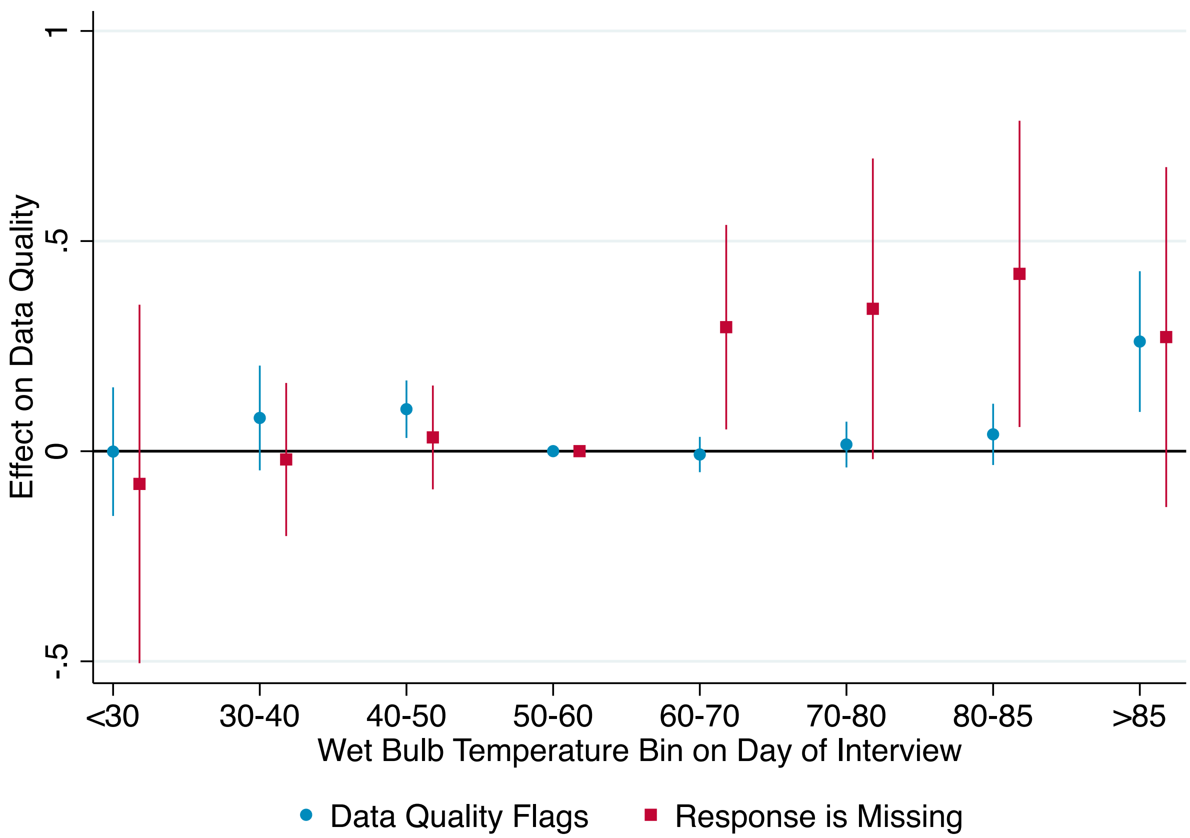
Note: This figure shows the results of interviewer-day level regressions using the number of interviews completed per hour worked as the outcome variable of interest, where hours worked is defined as the time between the start time of the first individual interview and the end time of the last individual interview in that day. The independent variables of interest are indicators for whether the daily average wet bulb temperature in the day of interview fell into the given bin. The regressions also include fixed effects for the survey round by region of country as well as controls for the characteristics of the set of respondents, the 10 year average of wet bulb temperature in the survey cluster in the month of interview, number of daylight hours, and the number of completed days in the survey round. Standard errors are clustered at the region-of-country level. Point estimates and 95% confidence intervals are shown.

Figure 2.6: Total Number of Interviews Completed in a Full Day Does Not Respond to Temperature



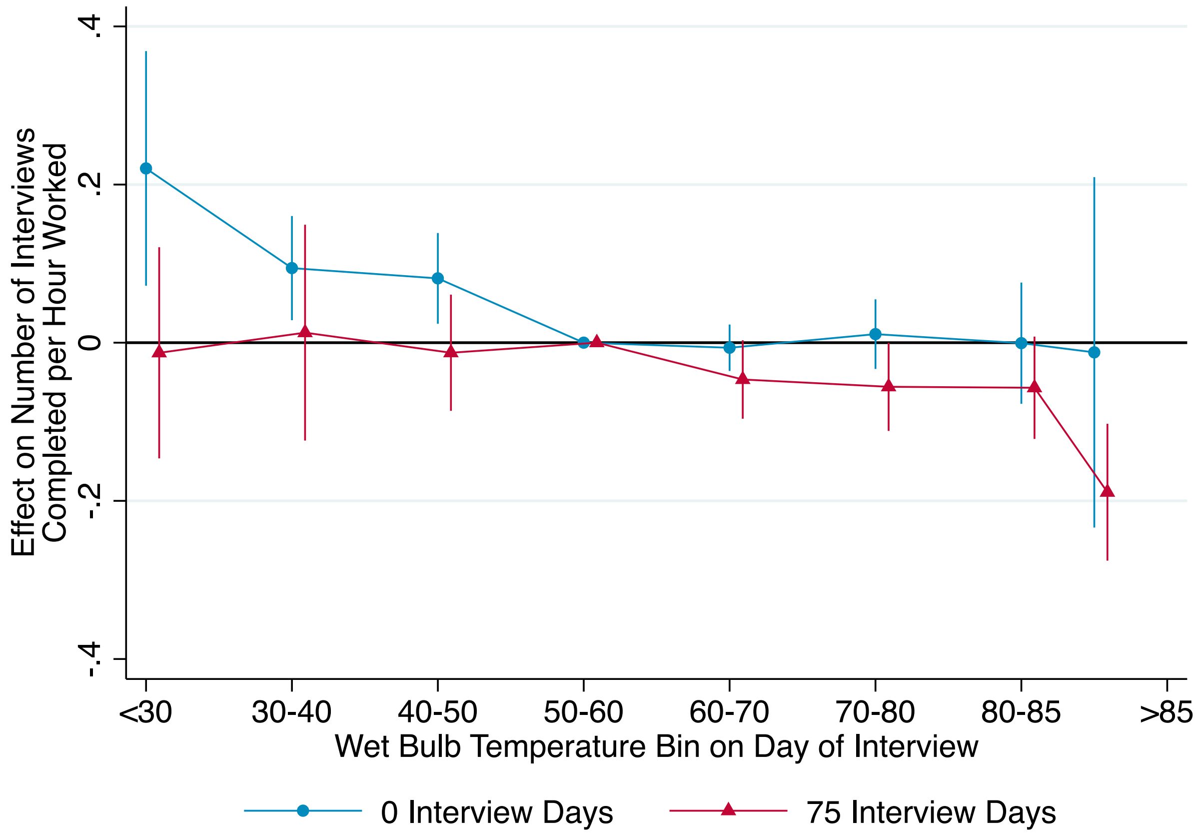
Note: This figure shows the results of interviewer-day level regressions using the number of interviews completed in a day as the outcome variable of interest. The independent variables of interest are indicators for whether the daily average wet bulb temperature in the day of interview fell into the given bin. The regressions also include fixed effects for the survey round by region of country as well as controls for the characteristics of the set of respondents, the 10 year average of wet bulb temperature in the survey cluster in the month of interview, number of daylight hours, and the number of completed days in the survey round. Standard errors are clustered at the region-of-country level. Point estimates and 95% confidence intervals are shown.

Figure 2.7: More Data Quality Problems Arise on Hot Days



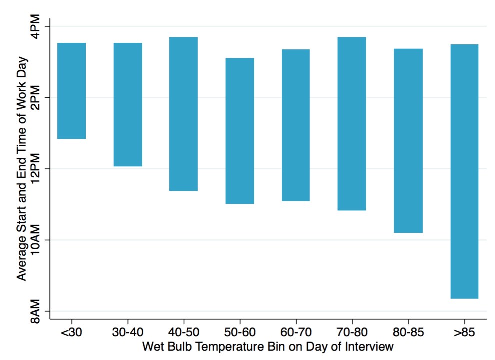
Note: This figure shows the results of interview-level regressions using counts of data quality problems as the outcome variables of interest. The independent variables of interest are indicators for whether the daily average wet bulb temperature in the day of interview fell into the given bin. The regressions also include fixed effects for the survey round by region of country as well as controls for respondent characteristics, the 10 year average of wet bulb temperature in the survey cluster in the month of interview, number of daylight hours, and the number of completed days in the survey round. Standard errors are clustered at the region-of-country level. Point estimates and 95% confidence intervals are shown.

Figure 2.8: Heterogeneity: The Effect of Heat on Number of Interviews Per Hour Increases with Experience

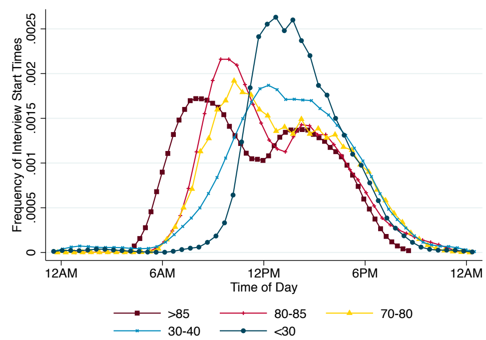


Note: This figure shows the results of interviewer-day level regressions using the number of interviews completed per hour worked as the outcome variable of interest, where hours worked is defined as the time between the start time of the first individual interview and the end time of the last individual interview in that day. The independent variables of interest are indicators for whether the daily average wet bulb temperature in the day of interview fell into the given bin. Each wet bulb bin is interacted with a measure of how many days the interviewer has worked on that survey round. The blue line shows the effect for an interviewer on his/her first day, and the red line shows the effect for the interviewer's 75th day. The interaction effects are statistically significant for each bin above 70 degrees, as well as for the ¡30 degree bin. The regressions also include fixed effects for the survey round by region of country as well as controls for the characteristics of the set of respondents, the 10 year average of wet bulb temperature in the survey cluster in the month of interview, number of daylight hours, and the number of completed days in the survey round. Standard errors are clustered at the region-of-country level. Point estimates and 95% confidence intervals are shown.

Figure 2.9: Mechanisms: Interviewers Begin Earlier on Warmer Days



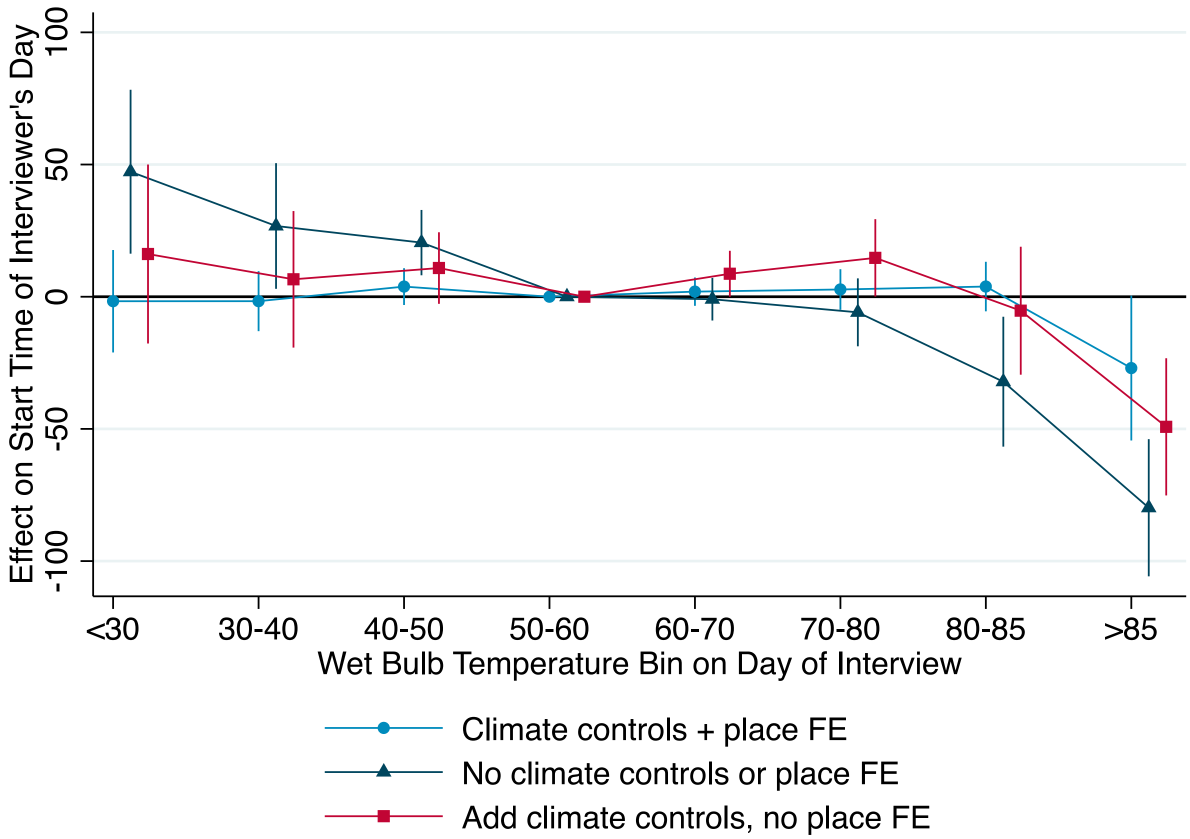
(a) Working Hours



(b) Distribution of Start Times

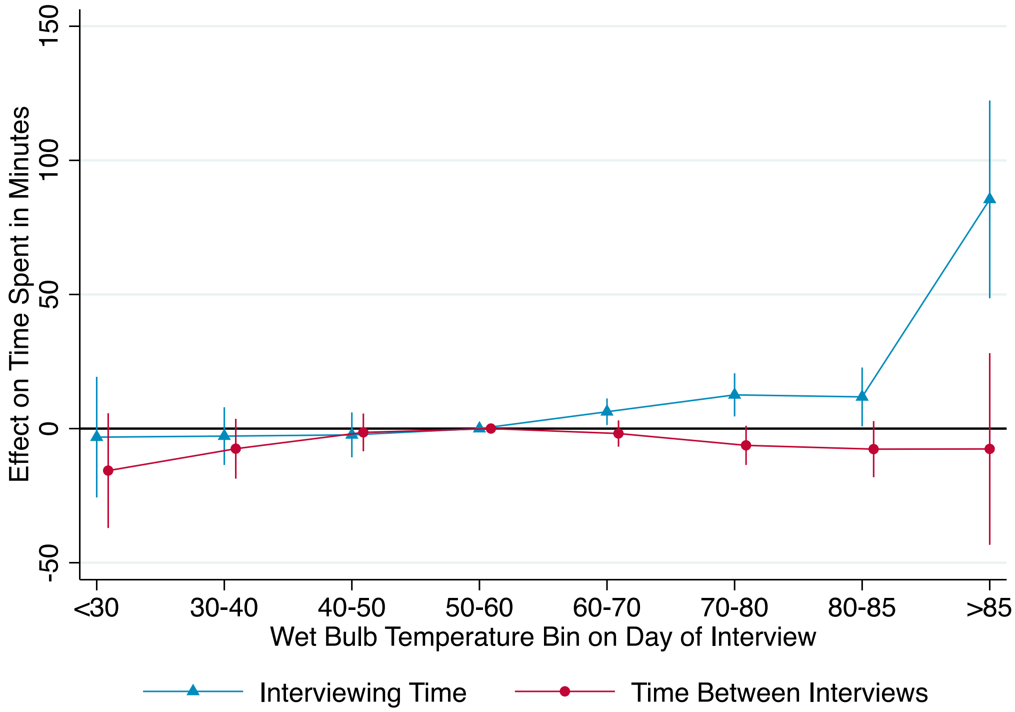
Note: Panel A shows summary statistics on the average start time of first individual interviews and end time of last individual interviews of interviewer-days in the sample by wet bulb temperature bins. Panel B shows the distribution of all interview start times throughout the day in the raw sample, broken into wet bulb bins.

Figure 2.10: Mechanisms: Change in Working Hours is Larger in Response to Usual Climate than to Surprise Weather Days



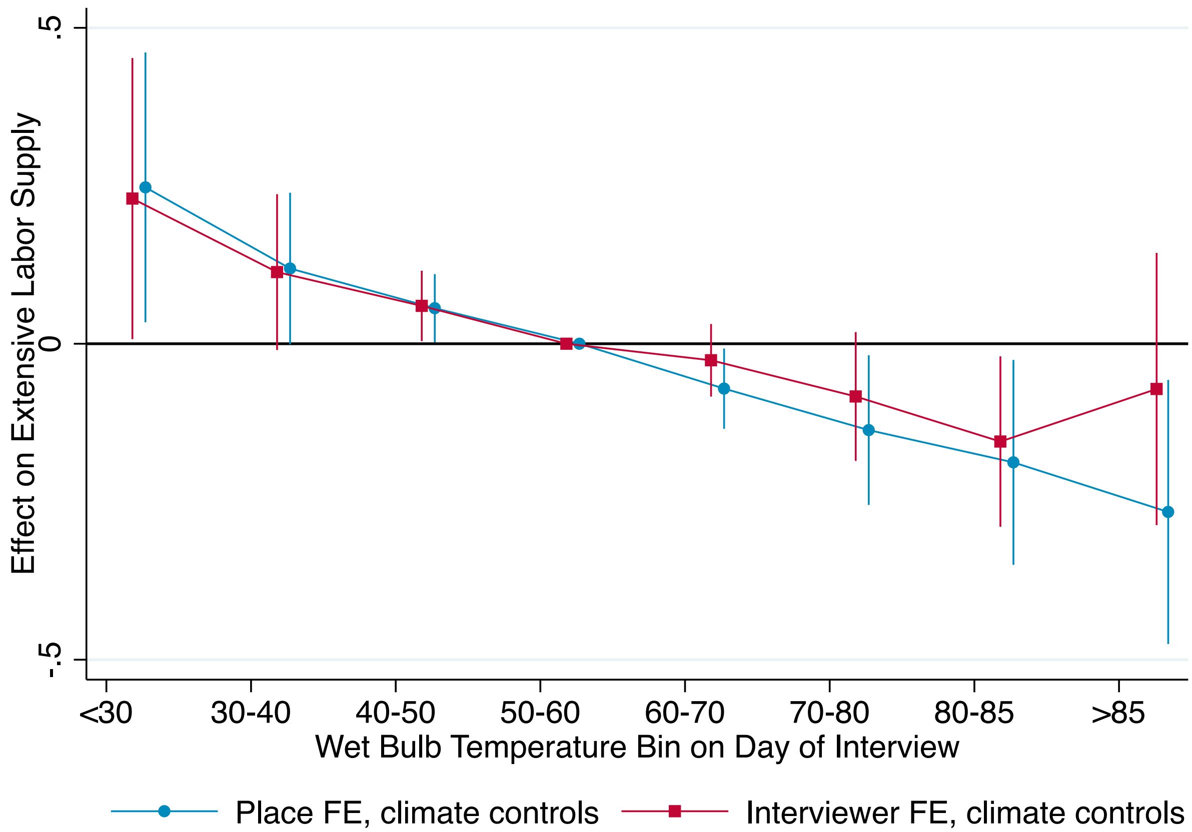
Note: This figure shows the results of interviewer-day level regressions using the start time of the interviewer-day as the outcome variable of interest. The independent variables of interest are indicators for whether the daily average wet bulb temperature in the day of interview fell into the given bin. Each line represents a different specification: the light blue line is the full specification with region of country by survey round fixed effects and controls for the 10-year average of wet bulb temperature in the month of interview in the survey cluster of interview. The red line removes the place fixed effects, and the dark blue line additionally removes the 10-year climate controls. All regressions include controls for the characteristics of the set of respondents, number of daylight hours, and the number of completed days in the survey round. Standard errors are clustered at the region-of-country level. Point estimates and 95% confidence intervals are shown.

Figure 2.11: Increase in Working Time Mainly Driven by Increase in Interviewing Time



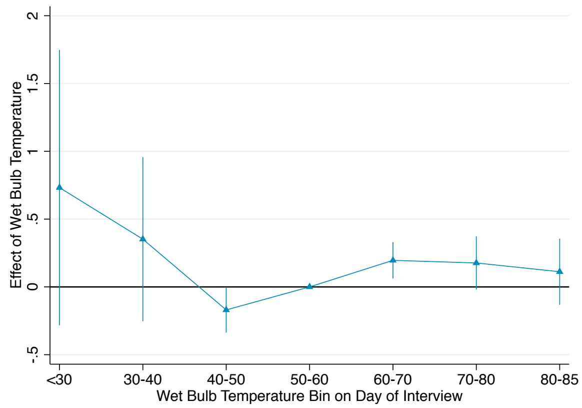
Note: This figure shows the results of interviewer-day level regressions using total interviewing time and total time between interviews, respectively, as the outcome variables of interest. The independent variables of interest are indicators for whether the daily average wet bulb temperature in the day of interview fell into the given bin. The regressions include region of country by survey round fixed effects, controls for the 10-year average of wet bulb temperature in the month of interview in the survey cluster of interview, controls for the characteristics of the set of respondents, number of daylight hours, and the number of completed days in the survey round. Standard errors are clustered at the region-of-country level. Point estimates and 95% confidence intervals are shown.

Figure 2.12: Effect on Labor Supply: Interviewers Less Likely to Conduct Interviews on Hot Days

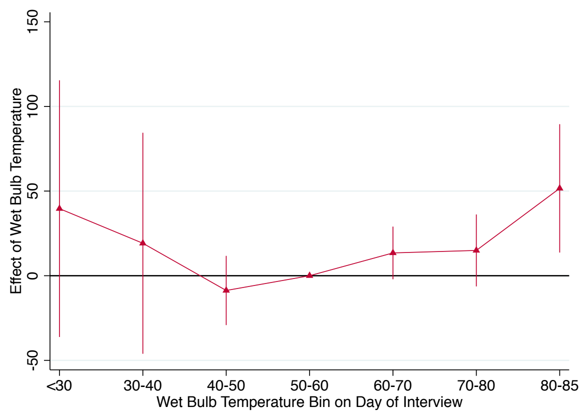


Note: This figure shows the results of interviewer-day level regressions using a dummy variable for whether the interviewer was observed conducting interviews that day as the outcome of interest. The independent variables of interest are indicators for whether the daily average wet bulb temperature in the most recent survey cluster visited by the interviewer on the day of interview fell into the given bin. The red line includes fixed effects for the unique interviewer while the blue line includes fixed effects for the region of country by survey round. Each regression additionally includes controls for the 10-year average of wet bulb temperature in the month of interview in the survey cluster of interview, the number of daylight hours, and the number of completed days in the survey round. Standard errors are clustered at the region-of-country level. Point estimates and 95% confidence intervals are shown.

Figure 2.13: Heterogeneity: Interviewers Conduct Fewer Interviews, Work Fewer Hours on Hot Days in Countries with Low Quality Public Sector Management



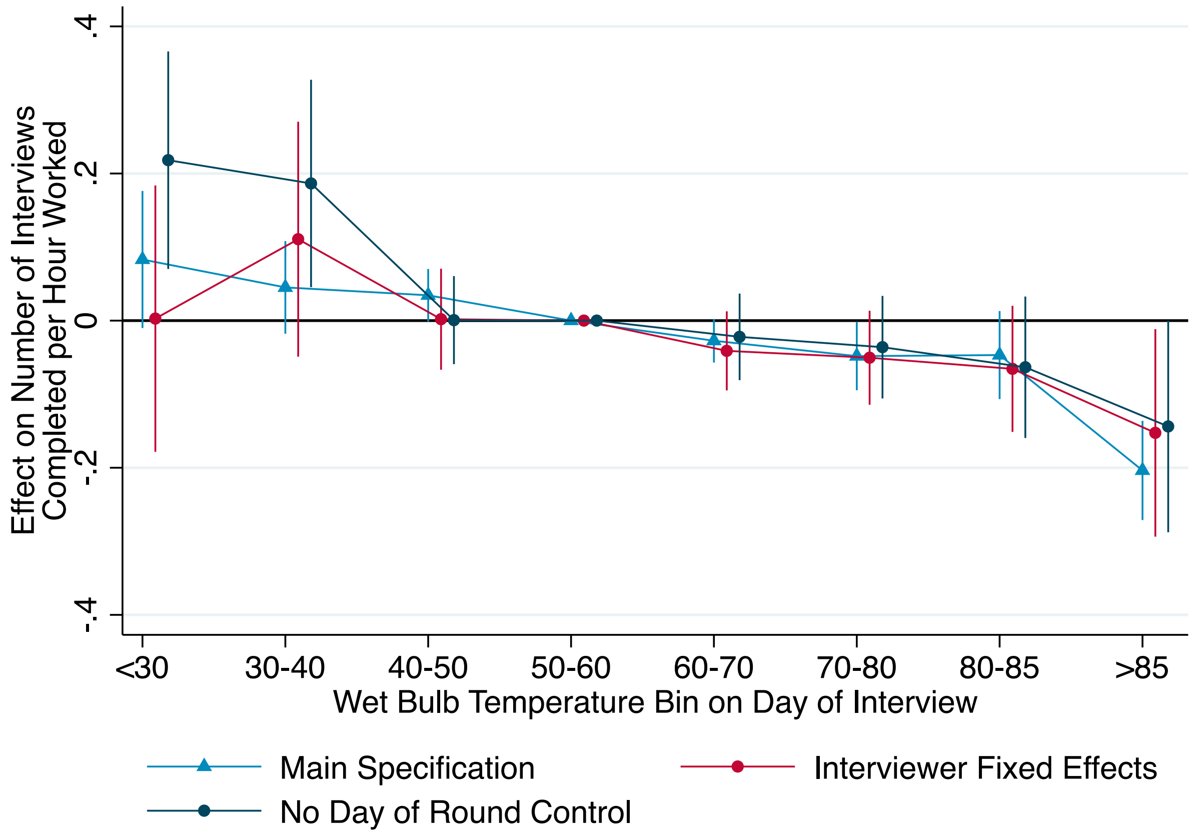
(a) Number of Interviews



(b) Working Time

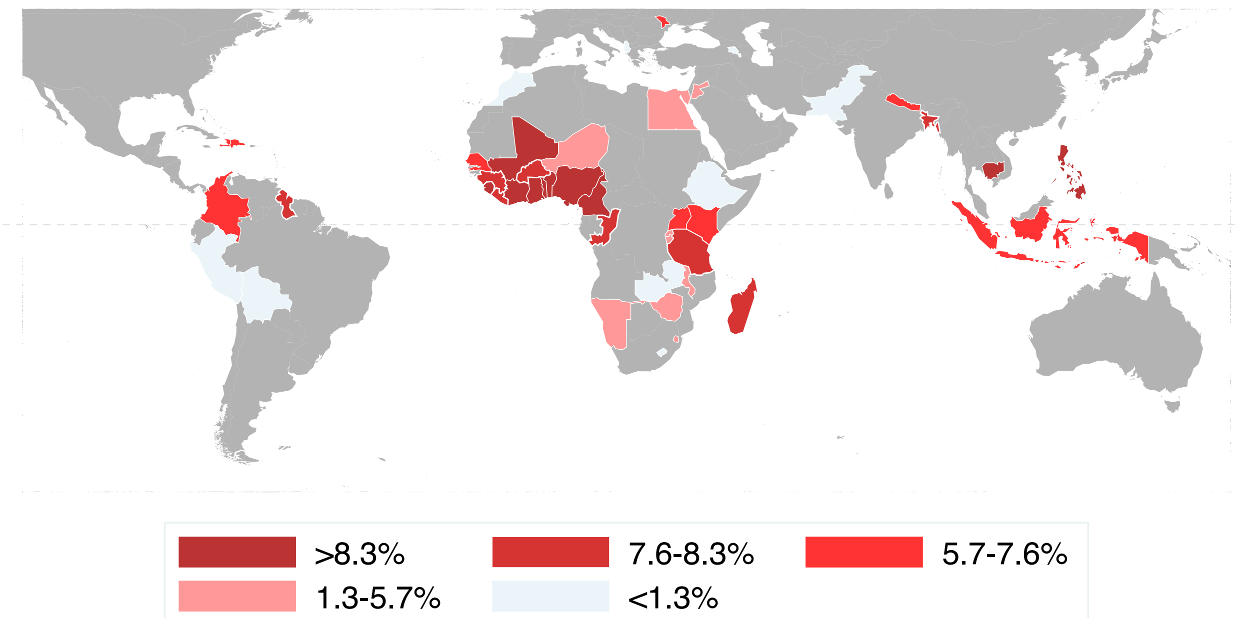
Note: This figure shows the result of regressions where the independent variables of interest are indicators for whether the daily average wet bulb temperature in the day of interview fell into a certain bin, interacted with a continuous measure of quality of public sector management from the World Bank Country Policy and Institutional Assessment. The regression is run at the level of the interviewer-day; Panel A shows number of interviews completed in a day as the outcome variable and Panel B shows the impact on minutes worked. The two figures plot the interaction effects, interpretable as the change in the outcome variable for a one unit increase in the public sector management index (a larger index value indicates better quality management). The regression also includes fixed effects for the survey round by region of country as well as controls for the characteristics of the set of respondents, the 10 year average of wet bulb temperature in the survey cluster in the month of interview, number of daylight hours, and the number of completed days in the survey round. Standard errors are clustered at the region-of-country level. Point estimates and 95% confidence intervals are shown.

Figure 2.14: Robustness: Results Robust to Alternative Controls, Fixed Effects



Note: This figure shows the results of interviewer-day level regressions using the number of interviews completed per hour worked as the outcome variable of interest, where hours worked is defined as the time between the start time of the first individual interview and the end time of the last individual interview in that day. The independent variables of interest are indicators for whether the daily average wet bulb temperature in the day of interview fell into the given bin. The light blue line gives the main specification, with fixed effects for the survey round by region of country and control variables for respondent characteristics, the 10 year average of wet bulb temperature in the survey cluster in the month of interview, number of daylight hours, and the number of completed days in the survey round. The red line replaces region of country fixed effects with interviewer fixed effects, while the dark blue line removes the day of round control variable. Standard errors are clustered at the region-of-country level. Point estimates and 95% confidence intervals are shown.

Figure 2.15: Implications: Median DHS Survey Would Require 5% Fewer Interviewer Hours with Mild Weather



Note: This figure visualizes the number of additional interviewer hours in the field that were required to complete the most recent DHS survey round in each country, relative to a counterfactual scenario in which all interviewing days had an average wet bulb temperature of 50-60 degrees. I calculate the actual number of interviews completed per hour worked by each interviewer on each day of the survey rounds, and then calculate what the number of interviews completed per hour would have been under the counterfactual scenario according to the estimates from Figure 2.5. I then calculate the implied difference in working time and add this measure up for each interviewer-day in the survey round before calculating this difference as a percent of actual time worked.

Table 2.1: Summary Statistics: Interview and Respondent Characteristics

	Mean	Standard Deviation
Respondent variables		
Age	29.258	9.716
Illiterate	0.347	0.476
Has Electricity	0.451	0.498
House Made of Formal Materials	0.549	0.498
Works	0.516	0.500
Data Should Exist, But Doesn't	0.972	4.289
Data Quality Flags	0.899	1.548
Interviewer/day variables		
Number of Interviews/Day	2.945	2.053
Hours Worked	4.497	3.357
Number of Interviews/hour	1.016	0.994

Note: This table contains summary statistics on major respondent control variables and outcome variables in the dataset. Observations are individual interviews for all except the last three outcome variables (number of interviews/day, hours worked, interviews/hour), where the observations are interviewer-days.

Table 2.2: Number of Interviews Completed Per Hour Worked Declines on Hot and Humid Days

Dependent variable:	Number of Interviews Completed Per Hour Worked			
	Dry Bulb	Dry Bulb X Humidity		Wet Bulb
		High Hum.	Low Hum.	
	(1)	(2)	(3)	(4)
>95	-0.090	-0.139**	-0.048	
	(0.105)	(0.062)	(0.097)	
90-95	-0.018	-0.101***	0.013	
	(0.035)	(0.039)	(0.033)	
85-90	-0.03	-0.091***	-0.005	-0.200***
	(0.026)	(0.028)	(0.025)	(0.034)
80-85	-0.017	-0.063***	0	-0.046
	(0.023)	(0.024)	(0.022)	(0.030)
70-80	-0.035	-0.068***	-0.016	-0.048**
	(0.022)	(0.024)	(0.020)	(0.024)
60-70	-0.017	-0.02	-0.011	-0.036**
	(0.018)	(0.026)	(0.018)	(0.016)
40-50	-0.009	-0.073**	0.006	0.035*
	(0.020)	(0.034)	(0.019)	(0.018)
30-40	-0.022	-0.05	-0.012	0.046
	(0.042)	(0.049)	(0.043)	(0.032)
<30	-0.024	-0.291***	-0.003	0.083*
	(0.045)	(0.051)	(0.047)	(0.047)
High Humidity		0.058***		
		(0.021)		
F-Stat: Weather	1.19	3.71		5.98
Observations	314,723	314,723		314,723
Region of country FE	X	X		X
Climate Controls	X	X		X

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Note: This table shows the results of interviewer-day level regressions using the number of interviews completed per hour worked as the outcome variable of interest, where hours worked is defined as the time between the start time of the first individual interview and the end time of the last individual interview in that day. Column 1 investigates the impact of dry bulb temperature, while columns 2 and 3 display the results from a single regression investigating effects of dry bulb temperature bins interacted with an indicator for high humidity, defined as the 75th percentile of humidity in the cluster of interview in the month of interview over 10 years. Finally, column 4 displays the effect of wet bulb temperature on the day of interview. The table also displays F-statistics for the coefficients on all contemporaneous temperature, temperature X humidity, and wet bulb bins. All regressions include controls for the set of respondents, the number of completed days in the survey round, number of daylight hours, the 10-year average of wet or dry bulb temperature in the survey cluster in the month of interview, and fixed effects for the survey round by region of country. Standard errors are clustered at the region-of-country level.

Table 2.3: Role of the Respondent: Some Selection on Respondent Characteristics

	Age	Works	Number of Children Aged <5	Formal House	Illiterate	Electricity	Married	Asset Index	Sex
Wet Bulb Bin									
<30 Degrees	0.411*	0.012	0.011	-0.011	-0.012	0.016	-0.002	0.034	-0.025*
	(0.249)	(0.028)	(0.041)	(0.027)	(0.022)	(0.025)	(0.016)	(0.121)	(0.015)
30-40 Degrees	0.065	-0.010	0.097**	-0.0529*	0.014	0.003	-0.005	-0.084	-0.006
	(0.163)	(0.017)	(0.049)	(0.030)	(0.019)	(0.027)	(0.010)	(0.134)	(0.007)
40-50 Degrees	0.009	-0.001	0.018	-0.033***	0.025***	0.006	-0.002	-0.012	-0.002
	(0.071)	(0.008)	(0.017)	(0.013)	(0.009)	(0.011)	(0.004)	(0.051)	(0.004)
60-70 Degrees	0.043	-0.002	-0.0257*	0.003	0.003	0.008	-0.001	-0.004	0.002
	(0.059)	(0.007)	(0.014)	(0.009)	(0.007)	(0.009)	(0.004)	(0.039)	(0.004)
70-80 Degrees	0.078	0.010	-0.010	0.000	-0.002	0.009	-0.001	0.011	0.002
	(0.095)	(0.010)	(0.021)	(0.013)	(0.009)	(0.013)	(0.006)	(0.055)	(0.006)
80-85 Degrees	0.026	0.012	0.013	-0.015	-0.001	0.005	0.003	0.017	0.010
	(0.120)	(0.012)	(0.029)	(0.018)	(0.011)	(0.018)	(0.009)	(0.081)	(0.007)
>85 Degrees	0.500*	-0.0848*	0.103	0.003	-0.009	0.063	0.039	0.421	-0.019
	(0.282)	(0.051)	(0.071)	(0.098)	(0.088)	(0.093)	(0.026)	(0.327)	(0.032)
Observations	962,620	962,620	962,620	962,620	962,620	962,620	962,620	962,620	1,221,676
R-squared	0.036	0.173	0.165	0.394	0.413	0.503	0.242	0.434	0.160

Clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: This figure shows the results of interview-level regressions using respondent characteristics as the outcome variables of interest. The independent variables of interest are indicators for whether the daily average wet bulb temperature in the day of interview fell into the given bin. The regressions also include fixed effects for the survey round by region of country as well as controls for the 10 year average of wet bulb temperature in the survey cluster in the month of interview, number of daylight hours, and the number of completed days in the survey round. Standard errors are clustered at the region-of-country level.

Chapter 3

The Consequences of Social Inequality for the Health and Development of India's Children: The Case of Caste, Sanitation, and Child Height[1]

3.1 Introduction

The relationship between socio-economic status and health is well established: across many societies, more advantaged groups are healthier, on average. White people in the United States can expect to live over three years longer than Black Americans, for example. Throughout the world, indigenous people have large gaps in life expectancy relative to other citizens: the gap is 20 years in Australia, 20 in Nepal, and 13 in Guatemala (UN, 2010). These differences in life expectancy and health status have been documented and studied extensively in the sociology and economics literatures. Beyond life expectancy, gaps in levels of diverse health outcomes, such as obesity (David W Johnston and Wang-Sheng Lee, 2011), and inputs, such as smoking (JY Ho and IT Elo, 2013), exist between racial groups. These gaps are present even in infancy: mortality rates among black infants are more than double those among white infants in the US. Racial gaps in infant mortality have persisted

[1] This paper was published in Social Justice Research, as Melissa LoPalo, Diane Coffey and Dean Spears (2019). This paper implemented a novel empirical test of a theory developed in prior work by Coffey and Spears. My contribution was to finalize the empirical analysis for this paper, conduct the literature review, write the main draft, develop the final empirical exhibits, and incorporate comments by referees in the publication process.

even in the context of declining overall risks of infant death (W Parker Frisbie, Robert A Hummer, Daniel A Powers, Seung-Eun Song and Starling G Pullum, 2010). Gaps in actual and perceived safety may also contribute to health inequities, both through direct effects of violence and through indirect effects on emotional well-being (Anne Case and Ta-Nehisi Coates, 2017).

The mechanisms behind these disparities are difficult to understand, as the effects of social rank and of economic deprivation are closely linked. Social group and income are highly correlated, so reasons for disparities between rich and poor people or people of different races can encompass both explanations. Similarly, places with high inequality are often places where many people are poor, so what appears to be an effect of inequality may actually simply reflect poverty. Finally, poor health can cause poverty, and variables such as education affect health as well as socioeconomic status. The many levels of causation and feedback loops inherent in the relationships among economic standing, poverty, and social standing make it difficult to draw definitive conclusions about the causal role of social status in health outcomes (David M. Cutler, Adriana Lleras-Muney and Tom Vogl, 2011).

Despite these difficulties, there has been substantial effort to examine the relative contributions of poverty and inequality to observed gaps. Research on occupational rank, social status, and health has sought to establish a substantial role of relative social standing or economic status. One classic example of the difficulty is the Whitehall Studies of British civil servants. Participants with lower status jobs had higher mortality rates, higher obesity rates, and less healthy behaviors, a fact which has been interpreted as evidence of an effect of status on health (Michael G Marmot, Geoffrey Rose, Martin Shipley and Peter J Hamilton (1978); Michael G Marmot, Stephen Stansfeld, Chandra Patel, Fiona North, Jenny Head, Ian White, Eric Brunner, Amanda Feeney and G Davey Smith (1991)). Yet, when Anne Case and Christina Paxson (2008) re-analyzed the Whitehall data with attention to effects over the life course, they found evidence that these correlations in important part reflect an effect of health—especially early-life health—on economic outcomes.

The mechanisms that these studies examine are direct in that they constitute an impact of individual-level characteristics on a person's health outcomes. But what if social inequality affects the health outcomes of all members of society? If so, such a finding about social inequality may escape the common threat that apparent evidence of economic inequality may simply reflect poverty (Diane Coffey and Dean Spears, 2017). In economics, cases where one person's decisions affect another person's outcomes are called externalities. One classic example of a negative externality is pollution, where a firm's production decisions create a byproduct that harms people living nearby. Externalities can also stem from individual decisions, such as the decision to vaccinate one's children, which protects not only that child, but also other children interacting with him or her. Vaccination is an example of a positive externality. A recent literature in development economics finds that in certain contexts, the social mechanisms that perpetuate and result from inequalities among groups can also cause health behaviors harmful to everyone, rather than just the disadvantaged group.

This paper introduces the reader to this literature, particularly in the context of India, where health outcomes, compared with other developing countries, are worse than would be predicted based on average income. Children in India are shorter and more likely to die at the beginning of life than children in other countries with similar levels of economic development. Because one in five births this year will be in India, understanding early-life health in India is important for everyone concerned with the global distribution of human well-being in our times.

Several recent studies have attempted to explain India's disadvantages in early-life health, finding that various dimensions of India's hierarchical social systems offer potential explanations for its relatively poor health outcomes. For example, in India, women are at their lowest social position within the household at the same time that they are most likely to have children. This low social status is associated with restricted food intake and heavy manual labor. Diane Coffey (2015) uses data from 2005 to find that 42.2 percent of women are underweight prior to becoming pregnant; this is a much higher level of pre-pregnancy underweight than in Sub-Saharan Africa, a region which is poorer. Because underweight mothers often give birth to low-birth weight babies, this research suggests that not only does the low social status ascribed to young women in India affect the health of that group; it also affects the birth outcomes of their children. This is of particular importance given the recent research linking child health with adult outcomes such as wages and educational attainment.

In this paper, we examine the effects of another source of social inequality in India on health outcomes: the caste system. We present new observational evidence on the role of casteism in the stunted growth of Indian children. The empirical exercise is based on a novel question asked in the India Human Development Survey (IHDS) in 2012, which asks whether respondents' households practice untouchability; that is, whether they enforce the caste hierarchy in their interactions with people from the lowest castes (Amit Thorat and Omkar Joshi, 2015). Substantial proportions of respondents answered in the affirmative, but there is large variation across regions and villages in the share of respondents who report that someone in their household practices untouchability. We examine the relationship between local casteism and child height, finding that children in households that practice untouchability are taller, on average, than children in households that do not practice untouchability, but that children living in villages with a larger share of residents practicing untouchability are shorter, on average.

The latter association is explained in large part by a third variable: open defecation. Our analysis provides evidence of an effect of local open defecation externalities on child height that derives from variation in the practice of untouchability in India. This exercise builds upon previous work by Coffey and Spears (2017), who explore the relationships between the norms of ritual purity and pollution that enforce caste boundaries and the practice of open defecation, as well as Dean Spears and Amit Thorat (2017), who show empirically that local levels of casteism predict open defecation behavior, even controlling for possible confounding variables. Finally, we discuss how this case sheds light on the overall relationship between social inequality and health outcomes.

3.2 Open Defecation, Health, and Height in India

This paper examines the link between the caste system, sanitation behavior, and population health in India. Sanitation in India is poor: about 55 percent of rural households defecate in the open (Diane Coffey and Dean Spears, 2018), and declines in open defecation have happened much more slowly in rural India than in the neighboring countries of Nepal and Bangladesh (Arabinda Ghosh, Aashish Gupta and Dean Spears, 2014). Furthermore, 55% is likely an underestimate of the true proportion of people who defecate in the open: the data come from the 2015-2016 National Family Health Survey, which asks about sanitation at the household level. Diane Coffey, Aashish Gupta, Payal Hathi, Nidhi Khurana, Dean

Spears, Nikhil Srivastav and Sangita Vyas (2014) find that in many households with latrines, some members use the latrine while others defecate in the open. Therefore, estimates of open defecation derived from household-level data will miss individuals who practice open defecation living in households in which others use a latrine. As in the case of maternal malnutrition, India is an outlier in sanitation: many countries that are much poorer than India have lower rates of open defecation. In fact, according to the World Health Organization and the United Nations Children's Fund Joint Monitoring Report (2017), rural Indian open defecation accounts for over half of the world's open defecation.

A large literature links the practice of open defecation to a plethora of health threats, especially for children. Unsafe disposal of feces spreads disease, and the consequences are not limited to those who practice open defecation themselves. Open defecation creates an infectious disease environment that threatens to harm the health and net nutrition of children during the critical early-life years. Previous research has tied local open defecation rates to infant mortality rates. Leveraging the fact that Hindus in India are more likely to defecate in the open than Muslims in India, Michael Geruso and Dean Spears (2018*b*) show that infant mortality rates are higher in villages with a larger fraction of Hindus (relative to Muslims), but that a household's own religion does not predict infant mortality, once neighborhood open defecation rates are accounted for. The overall Muslim survival advantage is striking because it exists despite the fact that on average, Muslims are disadvantaged in India relative to Hindus.

Children past infancy remain vulnerable during the critical early years in which their bodies and brains are developing: there are several potential pathways through which exposure to fecal germs can harm the health and proper growth of children. Most obviously, feces harbor germs that contribute to diarrheal episodes. Diarrhea itself can be dangerous, but even if it is only a temporary problem, it prevents the body from absorbing nutrition from food. Recent research has also suggested that long-term exposure to contaminated food and water may contribute to a condition called environmental enteric dysfunction, an illness that causes chronic intestinal inflammation. The inflammation over time causes a flattening in the folds of the intestine and prevents the proper absorption of food, which in turn has been hypothesized to stunt growth and cause other health problems (Jean H Humphrey, 2009). Feces also harbor parasites such as hookworms, which grow in children's intestines and siphon nutritional resources from the child, stunting growth.

These consequences for health carry far into the future; failure to grow to one's full height potential is correlated with failure to grow to one's full cognitive potential (Case and Paxson, 2008). This has consequences for economic outcomes such as educational attainment and earnings (Nicholas Lawson and Dean Spears, 2016). Given the correlations among height, overall health status, and cognitive ability, average population-level height is an important outcome to understand.

3.3 Open Defecation and the Caste System

This section briefly summarizes Coffey and Spears' (2017) recent book on the contribution of India's social systems to its unique open defecation problem. The caste system, historically and today, divides Indian society into many subgroups, called jatis. One's occupation, marriage prospects, and level of education are often related to one's caste, which is determined at birth. The caste system is particularly associated with Hinduism in India, although it is found among other religious groups as well.

The caste system is maintained and enforced in part through norms surrounding purity and pollution. The ideas of ritual purity and pollution are distinct from physical cleanliness and dirtiness. Under the norms of purity and pollution, both objects and people can be considered polluted and polluting to others or pure and purifying to others, irrespective of physical dirtiness. For example, cow dung and urine are considered purifying, while newborn babies and postpartum mothers are considered temporarily polluting.

The logic of purity and pollution reinforces the hierarchies of the caste system. People from the lowest-ranking castes are considered "untouchable" to other members of society. Higher-caste individuals have traditionally avoided physical contact with people from untouchable castes and refused to share food and water with them in order to avoid becoming polluted by them. People from untouchable castes are expected to do the dirtiest jobs, including disposing of animal carcasses, cleaning human feces, and de-blocking sewers and drains.

Initial improvements in sanitation, away from open defecation, have been achieved elsewhere in the developing world not through the installment of flush toilets but through the use of pit latrines, which are less expensive and simpler to install than sewers. These latrines, however, have pits that must eventually be emptied by hand, a job that, in India, is considered exceptionally polluting and thus only fit to be undertaken by people from untouchable castes. In recent years, people from these disadvantaged castes have protested the social injustices that they face in part by refusing to perform these types of jobs.

One consequence of these slow but ongoing social changes is that high-caste households considering whether or not to install a pit latrine are concerned that it will be difficult or expensive to find someone who is willing to empty the pit. As a result, many households forego the use of latrines altogether—even when the government distributes them for free.

In these ways the forces of social inequality in India have prevented the progress towards improved sanitation that have been achieved elsewhere in the developing world, including in much poorer places. Neighboring Bangladesh, for example, is poorer than India, on average, but has a culturally and religiously distinct society, which contributes to the fact that the use of inexpensive latrines is widespread and open defecation has essentially been eliminated there.

In this paper, we use quantitative data inspired by this qualitative evidence. The data are introduced in detail in the next section. A key variable in our analysis is the fraction of households in a village who report practicing untouchability—meaning, who report enforcing the rules of casteism. We use this variable as a proxy for the local importance of Hinduism-related casteism, purity, and pollution. Figure 1 shows that it is highly predictive of local open defecation. Therefore, our hypothesis is that local Hinduism-related social inequality, as proxied by reported practice of untouchability, causes children to be exposed to more open defecation, on average, which translates into reduced child height-for-age, a key marker of early-life health. The next section details the data, variables, and empirical methods by which we investigate this hypothesis.

3.4 Data and Methods

Building upon Coffey and Spears (2017), we pursue a novel empirical quantification of the impact of casteism on child health in India through the mechanism of open defecation. Because open defecation creates a disease environment that impacts entire localities, casteism could create substantial negative externalities through this pathway. We use the 2012 India

Human Development Survey (IHDS), which measured children's heights and asked about sanitation and casteism in a nationally representative sample of about 40,000 households. The IHDS is a panel survey implemented by the University of Maryland and the National Council of Applied Economic Research.

3.4.1 Dependent variable

The dependent variable in our analysis is a child's height-for-age z-score. We use the 2006 World Health Organization's child growth reference standards for healthy children to compute height-for-age z-scores. The z-score can be interpreted as the number of standard deviations away from the age-specific mean that a child's height falls.

Although the IHDS measured the height of some children over the age of 5, we restrict our sample to children under age 5 for comparability to studies using data from the Demographic and Health Surveys (DHS). We also follow the DHS recommendation of restricting the analysis to children with recorded height-for-age z-scores between -6 and 6. We highlight that the measurement of height in the IHDS is not as high quality as in the DHS: for Indian children in a similar time period, the variance of height in the IHDS is larger, and age-in-months is not available for all children. The calculation of height-for-age z score using WHO standards requires age in months. As a result, we approximate age-in-months as the midpoint based on age-in-years (so, a one-year-old would be coded as 18 months); this will introduce significant measurement error into the dependent variable. Therefore, we do not interpret any result in this section as a quantitatively precise estimate of India's true average effect of neighborhood open defection; rather, this section provides supporting evidence that suggests that such an effect exists and is large.

3.4.2 Explanatory variables

For the main explanatory variable of interest, we take advantage of new questions introduced in the 2012 wave of the IHDS. The survey sought to elicit attitudes regarding members of lower castes through a set of two questions. The first question was: "In your household do some members practice untouchability?" If the respondent answered in the negative, they were then asked "Would there be a problem if someone who is scheduled caste were to enter your kitchen or share utensils?" The enforcement of untouchability unfortunately remains common: about a quarter of households in rural India responded in the affirmative to the first question and 31 percent answered yes to the first or second question (Spears and Thorat, 2017). We follow Spears and Thorat in defining the variable of interest as an indicator for whether the respondent said yes to either question (see Figure 1 above).

The IHDS asks whether the household owns a toilet; the question also asks respondents specifically what type of latrine/toilet they own. The last option is "No facility belonging to household (or open fields)." Spears and Thorat (2017) use this answer as an indication of open defecation, and we follow that use. It is important to note, however, that Coffey et al. (2014) find that a substantial fraction of people living in households that own latrines do not use them, so this variable is likely to yield an undercount of actual open defecation. The two independent variables of interest (casteism and open defecation) are constructed as averages of the responses given by households living in the child's primary sampling units (PSU). These local averages estimate the fraction of a child's neighbors who practice open defecation and who practice untouchability; the child's own household is excluded from the local averages is both cases. PSUs in the IHDS are villages or urban blocks (IHDS 2011).

To test alternative mechanisms for the relationship between child height and PSU-level untouchability, we also include controls for levels of conflict in the village using two measures. The first gives a PSU-level average of respondent's opinion of the level of conflict in the village, on a scale from 1 to 3 (where 3 is the least conflict). The second gives a PSU-level average of whether respondents report that their community bonds together to solve local problem, as opposed to each family solving their problems individually.

Table 1 reports summary statistics for the full sample of children, and for subsamples of children exposed to different levels of local untouchability. Many important covariates, such as household consumption per capita, maternal education, markers for women's status (men eat first as opposed to household eating together or women eating first, a count of decisions that the mother reports having a say over out of eight total types of decisions), whether the household is urban, whether the household owns land, and number of children, are non-monotonically related to the prevalence of untouchability among a child's neighbors.

3.5 Empirical Strategy

Our observational empirical strategy builds upon Spears and Thorat (2017), who relate local levels of casteism to local open defecation. We use the same variable for casteism as they use as their independent variable—namely, the fraction of households in a survey cluster or primary sampling unit (PSU) who report practicing untouchability—to examine the links between casteism, open defecation, and child height. Spears and Thorat show that there is substantial variation in local practices of both open defecation and casteism across India. They then demonstrate that these two practices are correlated, with households that admit to enforcing untouchability being more likely to defecate in the open themselves and areas where more people practice untouchability having a larger fraction of people who practice open defecation. Although their estimates are not intended to uncover a causal effect of casteism on open defecation, they rule out several alternative explanations, such as associations between practicing untouchability and broader economic disadvantage as well as associations between the practice of untouchability and incorrect health beliefs. Their econometric analysis shows that this relationship is robust to a wide range of regression controls. Therefore, this study will use local practices of untouchability as measured by the 2012 IHDS question as a source of variation in open defecation to examine the consequences for population health. We build on Spears and Thorat, who do not study health or height, by taking the relationship that they document as the first stage for the full chain that we investigate in this paper.

This paper empirically investigates the link between casteism, open defecation, and height. We show that, although children living in households that report practicing untouchability are slightly taller than other Indian children, on average, children living in localities where more neighbors practice untouchability are shorter on average, whether or not a wide set of household socioeconomic observables are accounted for. The main analysis investigates the relationship between PSU untouchability, open defecation, and individual-level child height using reduced-form OLS regressions. Specifically, we investigate these relationships by first estimating equations of the following form:

$$HFA_{ihv} = \beta_1 practiceuntouchability_{hv} + \beta_2 \overline{practiceuntouchability}_v^{-i} + \beta_3 \overline{opendefecation}_v^{-i} + X_{ihv} + \epsilon_{ihv} \tag{3.1}$$

where i indexes individual children, h indexes households and v indexes PSU's (villages). $\overline{practiceuntouchability}_v^{-i}$ and $\overline{opendefecation}_v^{-i} + X_{ihv}$ are PSU averages of the indicators of household-level untouchability and open defecation, respectively, calculated without household h. X_{ihv} is a vector of child and household-level controls. This includes vectors of fully interacted age and sex controls, log consumption per capita of the household, whether the child was measured lying down or standing, whether the household owns land, indicator variables for the number of children in the house, and indicators for caste and religion categories. The standard errors are clustered at the level of the PSUs.

We examine the relationship between our independent variables of interest and height by building up to Equation 1 in stages. We first examine the relationship between a household's practice of untouchability and the height of children living in the household. We then successively add other independent variables, examining how the effect of village-level untouchability impacts height relative to own household untouchability, and finally examining whether measures of village-level open defecation mediate the relationship between village-level casteism and height outcomes.

This analysis shows that children living in localities where more neighbors report practicing untouchability are shorter, on average, but not after accounting for differences in exposure to open defecation. The quantitative, descriptive results from this exercise are consistent with open defecation having a large effect of child height in India, but we make no claim that these estimates reflect quasi-randomized causal effects.

3.6 Results

Table 2 shows the results of regressions examining the relationship between measures of untouchability at the household and PSU levels and child height-for-age. Column 1 shows the association between own household untouchability and height-for-age, with the full set of age-by-sex and measuring position controls. The relationship is small and statistically insignificant. Column 2 adds PSU-level untouchability to the regression and shows that children living among more neighbors who practice untouchability are shorter, on average, but those whose own households practice untouchability are taller. Column 3 adds a vector of controls for other predictors of child height, and the negative relationship between PSU untouchability and height remains. Note that controlling for PSU conflict and PSU bonding together could plausibly be over-controlling, because these could reflect common variation with untouchability; we nevertheless include column 3 for a more complete investigation of robustness. The positive relationship between own household untouchability and height shrinks with the inclusion of household observable characteristics, suggesting that this effect may be compositional rather than causal. Overall, these results demonstrate that the association between child height and untouchability are driven by neighbors' practice of untouchability, not the child's own household. This suggests that the negative overall relationship between the practice of untouchability and height partially reflects an externality of that practice on other households in the neighborhood.

Table 3 builds on Table 2 by investigating open defecation as a mediator of the relationships displayed in Table 2. The table presents three pairs of OLS results, each pair with and without a control for PSU-level open defecation included among the independent variables. All regressions control for age (in years) interacted with sex and whether the height measurement was taken standing or lying down. The first pair of estimates shows the results of a regression of height on PSU untouchability with and without the control for PSU-level open defecation. The second pair adds a control for urban status of the PSU and other PSU-level characteristics, and the third adds household-level controls for other predictors of child height.

The pairing exercise is useful because it helps us verify that open defecation is a major mechanism behind the relationship between untouchability and child height. There are various reasons why local casteism might predict child height (such as availability of public goods, or consequences of mothers' stress); by reporting each regression with and without the open defecation control, we demonstrate that the association between the PSU-level untouchability variable and open defecation explains a major portion of the negative relationship between untouchability and height. The evidence from Table 2 suggested that the effects of untouchability on height occurred through an effect on the neighbors of households that practice untouchability, rather than through an association with own-household discriminatory behavior. This evidence strengthens that argument by testing several alternative hypotheses.

In columns 3 and 4, we add in controls for several other PSU-level characteristics: urban status and controls for the level of conflict in the PSU. These controls account for PSU-level characteristics that may be correlated with both practices of untouchability and open defecation behaviors. For instance, it may be the case that untouchability is correlated with an overall lack of social cooperation or capital in the village, which could also be correlated with open defecation behavior. However, the overall pattern from columns 1 and 2 holds with these additional controls. Finally, in columns 5 and 6, we control for several socioeconomic and other variables of the child's household, including maternal education and women's status, which may be correlated with practices of untouchability and with child height.[2]

[2]In results not shown here but available upon request, we exclude Dalit respondents from the computation of each PSU's fraction of households reporting untouchability. We include this as a robustness check because

Although the negative relationship between untouchability and height is statistically insignificant in the specification without the open defecation control, accounting for open defecation dramatically decreases the magnitude of the relationship. The controls verify that this pattern of results is not merely due to differences between rural and urban India, nor to the socioeconomic status of the child's household, nor to the child's family's caste or religion category.

These results are a useful addition to the literature on the effects of open defecation on child height because household-level omitted variables are implausible: children in India who live in households that report practicing untouchability are taller on average, than children who live in households who do not. Therefore, the fact that children living among more local neighbors who report practicing untouchability are shorter on average, suggests a mechanism rooted in externalities and context effects.

The confidence intervals in this analysis are wide, and these quantifications should not be taken as more than qualitatively indicative of the importance of open defecation for child height. However, in each case the null hypothesis of no association between sanitation and child height can be rejected, and in each case the confidence interval includes the effect sizes large enough to explain 100% of the height gap between children in India and sub-Saharan Africa, as quantified by Dean Spears (2018). This is true with or without a wide set of controls.

---

a small, but positive number of respondents from untouchable castes report that they practice untouchability, which is difficult to interpret. The regression results are qualitatively unchanged, although this necessarily drops from the sample children living in segregated, all-untouchable caste PSUs.

3.7 Discussion and Conclusion

This paper has illustrated a link among social hierarchies, health behaviors, and health outcomes in India. In this case, the norms dividing people from untouchable castes from non-untouchables create health hazards that harm everyone, not just the socially and economically disadvantaged. Ideas about ritual purity and pollution have prevented progress towards improved sanitation in India, which would save lives and help people have healthy childhoods and reach their full economic potential. Open defecation in rural India contributes to an infectious disease environment that plagues Indian children with diarrhea, parasites, and other health problems. We find an association between PSU shares of people practicing untouchability and child height. This is despite the fact that, controlling for PSU untouchability, own household untouchability predicts improvements in child height. This suggests that the mechanism at play is a consequence of casteism that affects an entire neighborhood, rather than just the household practicing untouchability. We show that a major mechanism is open defecation: once PSU-level open defecation is controlled for, PSU untouchability no longer predicts child height.

Despite our results, important uncertainties and limitations remain. One limit is that, although the IHDS is a large survey, our statistical power is limited by the fact that not all households have children, and that we are studying village-level variation and must therefore cluster standard errors. The most notable limit is that this is an observational study of existing variation. The advantage is that we are able to study the full range of exposure to open defecation and casteism that exists throughout rural India. However, we do not have the benefits for causal inference of a randomized controlled trial. As a result, we cannot fully rule out all possible confounding factors that may have influenced our results.

A large literature in economics has debated whether economic inequality causes poor health, or whether the apparent correlation reflects omitted variables and spurious factors (Anne Case and Christina Paxson (2011); Angus Deaton (2013)). Our analysis differs from this debate because we study a case of social inequality: we show novel evidence of a case where social inequality leads to worse health outcomes throughout society. Previous literature in psychology, sociology, and economics has suggested that social inequality may have a substantial impact on health through mechanisms such as feelings of relative deprivation and restricted access to health care. However, these mechanisms are difficult to disentangle from the direct impact of absolute poverty on health as well as from third variables associated with both relative deprivation and health. The results of this paper suggest that—without necessarily offering evidence for or against channels of causation between one's own rank and one's own health—in certain important contexts, social inequality can create health externalities that affect whole populations, rather than only the marginalized themselves.

Table 3.1: Summary statistics from IHDS analysis

	Full sample	No untouchability	Intermediate	All untouchability
Height-for-age z-score	-2.21	-1.98	-2.26	-2.52
PSU open defecation	0.52	0.35	0.56	0.65
PSU untouchability	0.33	0	0.39	1
PSU conflict	2.44	2.62	2.40	2.44
PSU band together	0.74	0.74	0.74	0.74
Urban	0.28	0.43	0.25	0.19
ln(cons. per capita)	9.62	9.71	9.60	9.57
HH farm land	0.48	0.31	0.52	0.53
Is only child	0.18	0.24	0.16	0.15
HH number of children	2.82	2.48	2.90	2.91
Hindu	0.80	0.64	0.83	0.84
Muslim	0.15	0.24	0.13	0.14
Dalit	0.25	0.27	0.25	0.06
Men eat first	0.32	0.17	0.35	0.38
Resp. has say	6.24	6.32	6.22	6.67
Mom has no education	0.35	0.22	0.37	0.48
Mom has less than secondary education	0.51	0.59	0.49	0.40
Mom has secondary education or more	0.15	0.19	0.14	0.12

Note: Child-level sample of n=12,922 children, matching child-level height analysis from the IHDS. No untouchability and all untouchability are subsamples for which PSU untouchability equals 0 and 1, respectively, with intermediate the remaining observations. The no untouchability, intermediate, and all untouchability subsamples account for 19.18 percent, 79.47 percent, and 1.35 percent of the full sample, respectively..

Table 3.2: IHDS results are driven by local untouchability context, not own-household's practice of untouchability

	(1)	(2)	(3)
Dependent variable:		Height-for-age z-score	
Own HH untouchability	-0.0144	0.148**	0.108+
	(0.049)	(0.057)	(0.060)
PSU untouchability		-0.445***	-0.184+
		(0.105)	(0.108)
PSU conflict			0.039
			(0.056)
PSU bond together			0.066
			(0.092)
ln(cons. per capita)			0.154***
			(0.045)
Urban			0.168**
			(0.0575)
HH has farm land			0.0185
			(0.0471)
Men eat first			✓
Resp. has say			✓
3 maternal education groups			✓
			✓
Number of children			✓
7 social groups			✓
Age, sex, position	✓	✓	✓
n	11,094	11,094	10,226

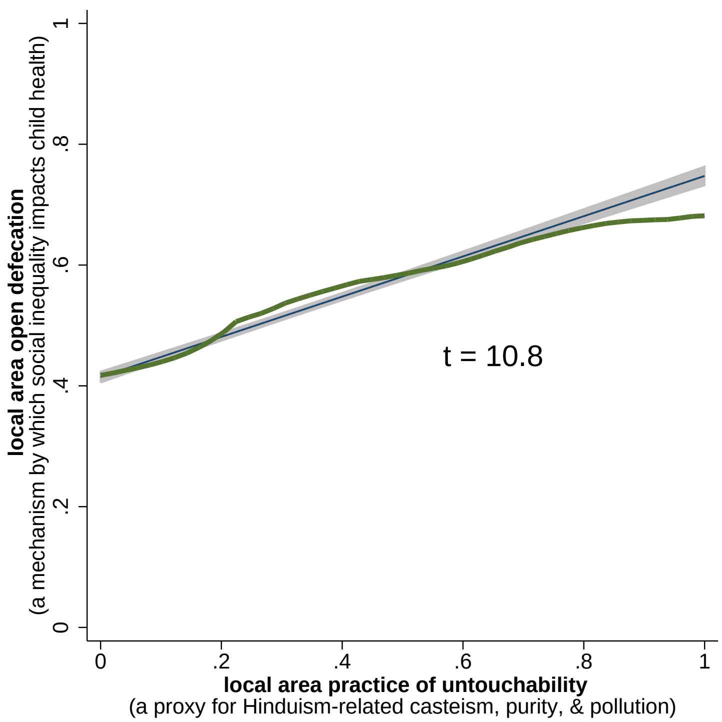
Note: Note: Child-level sample, matching child-level height analysis from IHDS.
***, **, *, and + denote statistical significance at the 0.001, 0.01, 0.05 and 0.1 levels. Standard errors are in parentheses.

Table 3.3: Evidence from the association between child height and the practice of untouchability, IHDS

	(1)	(2)	(3)	(4)	(5)	(6)
PSU untouch.	-0.296***	-0.118	-0.169+	-0.068	-0.083	-0.048
	(0.089)	(0.086)	(0.091)	(0.090)	(0.092)	(0.092)
PSU OD		-0.545***		-0.498***		-0.286***
		(0.047)		(0.052)		(0.056)
Urban			0.348***	0.131*	0.169**	0.091
			(0.053)	(0.059)	(0.058)	(0.061)
PSU conflict			0.086	0.088	0.038	0.047
			(0.058)	(0.059)	(0.056)	(0.056)
PSU band together			0.106	0.073	0.065	0.045
			(0.094)	(0.092)	(0.092)	(0.092)
ln (cons.)					0.155***	0.109*
per capita					(0.045)	(0.045)
HH farm land					0.021	0.027
					(0.047)	(0.047)
Maternal Education					✓	✓
Men eat first					✓	✓
Resp. has say					✓	✓
Number of children					✓	✓
7 social groups					✓	✓
Age, sex, position	✓	✓	✓	✓	✓	✓
N	11,094	11,094	11,094	11,094	10,226	10,226

Note: Note: Child-level sample from the India Human Development Survey 2012; children under 5 years old with height-for-age between -6 and 6. The dependent variable is height-for-age z-score. PSU open defecation is the fraction of interviewed households in the child's survey PSU that report defecating in the open, while "PSU untouch." is the fraction of households that report practicing untouchability. See the Data and Methods section for further details.
***, **, *, and + denote statistical significance at the 0.001, 0.01, 0.05 and 0.1 levels. Standard errors are in parentheses.

Figure 3.1: A proxy of Hinduism-related casteism, purity, and pollution practices predicts local-area exposure to open defecation among children



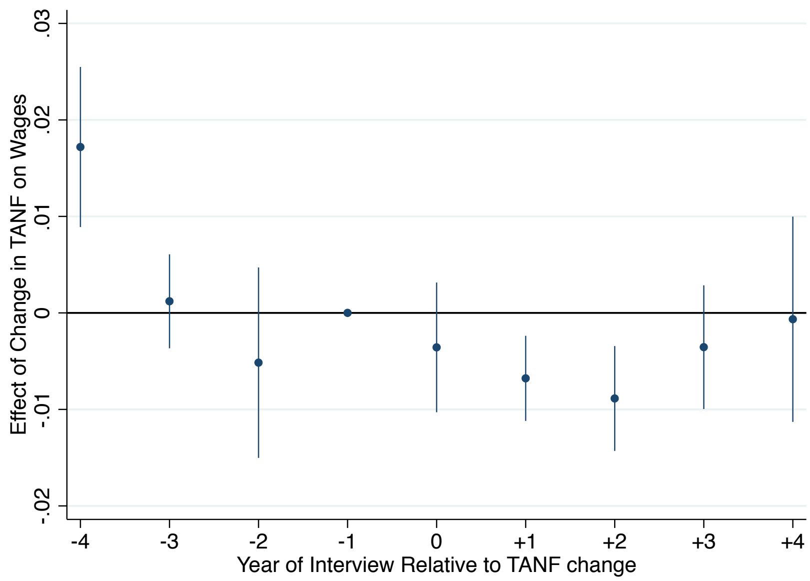
Note: Child-level sample of n=12,922 children, matching child-level height analysis from the IHDS. The horizontal axis displays the PSU average of our measure of untouchability and the horizontal axis displays the PSU average of open defecation.

Appendices

Appendix A

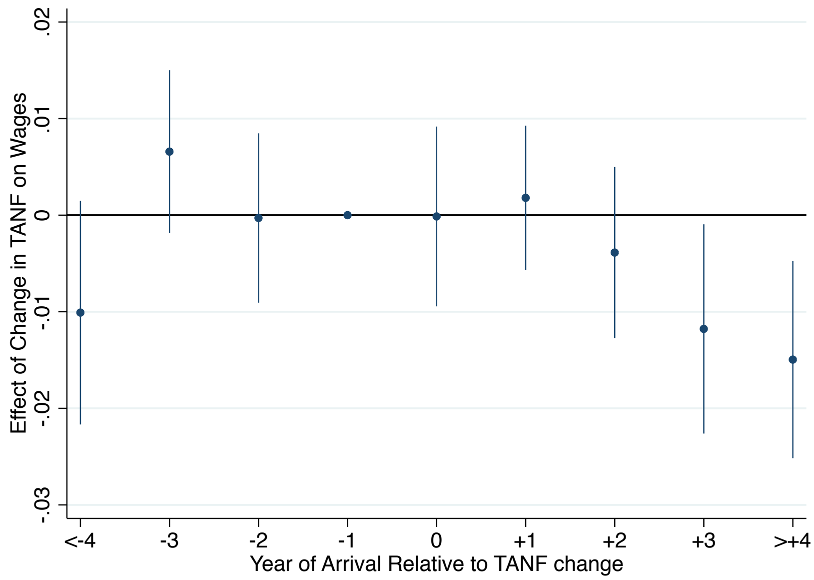
Appendix to Chapter 1

Figure A1: Event Study: Employed Native Wages Before and After Large TANF Increase



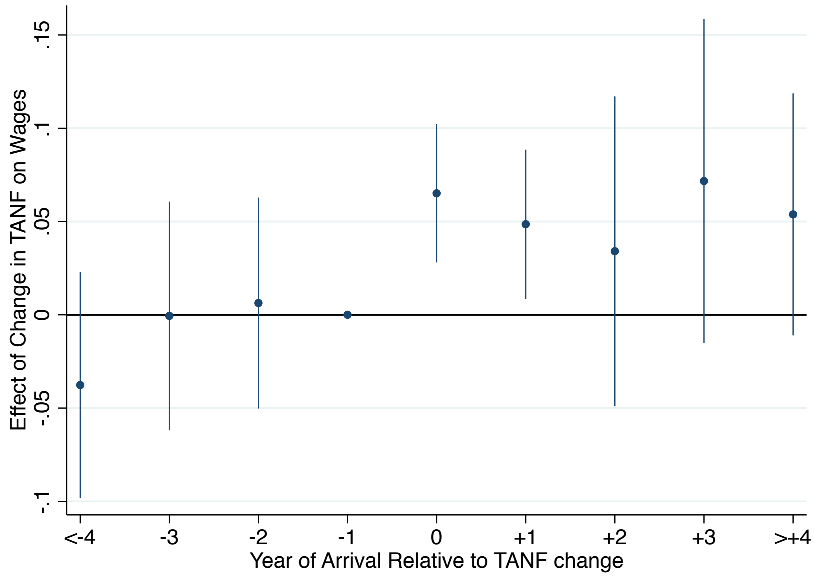
Note: This figure displays the results of a version of 1.1, where the independent variables of interest are indicator variables for whether a native was interviewed in the year of a large increase in TANF maximum benefit levels (at least 3 percent) as well as whether they were interviewed up to 4 years before or after such an increase. The dependent variable is log wages for native workers, excluding those with 0 wages. The categories of "4+" and "-4" also include natives who were interviewed greater than 4 years away from the increase in either direction. Natives in states that never experienced a large increase in TANF benefits are omitted in this sample. The regression also includes controls for marital status, sex, age and its square, number of own children in the household, and unemployment rate in the the year of interview. There are also controls for year of interview, and state fixed effects as well as state time trends. The standard errors are clustered at the state level.

Figure A2: Event Study: Non-Refugee Foreign Born Wages Before and After Large TANF Increase



Note: This figure displays the results of a version of Equation 1.1, where the independent variables of interest are indicator variables for whether an immigrant arrived in the year of a large increase in TANF maximum benefit levels (at least 3 percent) as well as whether they arrived up to 4 years before or after such an increase. The regression is run on the sample of non-refugee foreign-born individuals, excluding those with 0 earnings. The categories of "4+" and "-4" also include refugees who arrived greater than 4 years away from the increase in either direction. Immigrants in states that never experienced a large increase in TANF benefits are omitted in this sample. The regression also includes controls for marital status, sex, age and its square, number of own children in the household, unemployment rate in the year of arrival and the year of interview, and dummy variables for each of 11 sending regions. These regions are Mexico, Canada, Latin America (excluding Mexico), Northern and Western Europe, Eastern Europe, East Asia, Southeast Asia, Southwest Asia, the Middle East, Africa, and Oceania. There are also controls for year of arrival, year of interview, and state fixed effects as well as state time trends. The standard errors are clustered at the state level.

Figure A3: Event Study: Refugee Wages Before and After Large TANF Increase



Note: This figure displays the results of a version of Equation 1.1, where the independent variables of interest are indicator variables for whether a refugee arrived in the year of a large increase in TANF maximum benefit levels (at least 3 percent) as well as whether they arrived up to 4 years before or after such an increase. The categories of "4+" and "-4" also include refugees who arrived greater than 4 years away from the increase in either direction. Refugees in states that never experienced a large increase in TANF benefits are included in this sample, in the omitted category of "-1." The regression excludes refugees with 0 earnings. The regression also includes controls for marital status, sex, age and its square, number of own children in the household, unemployment rate in the year of arrival and the year of interview, and dummy variables for each of 14 sending countries, defined as in Table A2. There are also controls for year of arrival, year of interview, and state fixed effects as well as state time trends. The standard errors are clustered at the state level.

Table A1: Summary Statistics: Refugees and Economic Immigrants

	2000		2012-2016	
	Non-Refugee FB	Refugees	Non-Refugee FB	Refugees
Age	36.771	41.061	40.895	45.042
Marital Status	0.663	0.74	0.644	0.594
Share Female	0.493	0.504	0.518	0.51
Real Annual Wage	26,295.956	19,498.897	29,047.746	20,435.785
Number of own children	0.843	1.092	1.158	1.101
Years in the United States	2.155	2.564	10.094	9.514
English Ability	0.632	0.449	0.641	0.516
Employed	0.572	0.547	0.684	0.639
Labor Force Status	0.616	0.607	0.729	0.697
Urban Status	0.935	0.972	0.935	0.965
Enrolled in School	0.144	0.137	0.073	0.066
Years of Education	13.189	13.297	12.31	11.931
Migrated between States, Prior Year			0.025	0.021
Observations	114,051	9,419	628,085	34,567

Source: 2000 Census; 2012-2016 American Community Survey. A list of sending countries that comprise each sending region and their respective sample sizes is available in Table A2.
Note: This table gives the means of major control variables by refugee status and time period for the sample of refugees aged 25 and older.

Table A2: Refugee Sending Populations

Country	Average Refugee Share	Years in Sample	Total Sample Size
Former Soviet Union	0.732	1996-1998	17,017
Bosnia and Herzegovina	0.933	1996-2003	9,016
Croatia	0.648	2000-2002	388
Bhutan	0.940	2008-2016	1,237
Burma	0.893	2006-2016	2,502
Laos	0.744	2005-2006	271
Afghanistan	0.732	2001-2006; 2012-2016	938
Iraq	0.803	1996-1999; 2008-2016	5,043
Cuba	0.821	Full Sample	34,485
Congo	0.640	2006; 2008-2009; 2015-2016	100
Eritrea	0.675	2009-2016	372
Liberia	0.699	1999-2006	1,789
Somalia	0.877	Full Sample	3,522
Sudan	0.695	1999-2006	1,449

Note: This table shows which countries and years of arrival were coded as refugees in my sample. I made the selections based on Yearbook of Immigration Statistics data, coding a country as a refugee country if the fraction of refugee arrivals to immigrant visas issued plus refugee arrivals in the year the immigrant arrived exceeded 60 percent. The sample size column gives the number of refugees aged 25 and older observed in my sample. The rightmost column gives the average number of refugee arrivals over the years that the country is in the sample and the average number of immigrant visas issued to immigrants from the same country (this includes immediate relatives, family preference, employment preference, and diversity visas). The visa statistics come from the Annual Report of the Visa Office. Cuban Entrants as well as Special Immigrant Visa recipients are included as refugees, as they are eligible for ORR benefits. The refugee share from Bhutan exceeds 60 percent from 2008-2014, but the ACS does not begin identifying Bhutan as an origin country until 2012. The horizontal lines designate regions of origin.

Table A3: Instrumental Variables Results: Effects of TANF Generosity on Wages

Dependent Variable:	Log Wage			
	(1)	(2)	(3)	(4)
TANF Max Benefit	0.07	0.149***	0.094**	0.133***
	(0.046)	(0.031)	(0.040)	(0.025)
< High School	-0.380***	-0.453***	-0.380***	-0.453***
	(0.020)	(0.031)	(0.021)	(0.032)
Marital Status	-0.302***	-0.361***	-0.302***	-0.361***
	(0.012)	(0.021)	(0.012)	(0.022)
Female	-0.249***	-0.295***	-0.249***	-0.295***
	(0.039)	(0.046)	(0.039)	(0.047)
Age	0.227***	0.216***	0.227***	0.215***
	(0.012)	(0.013)	(0.012)	(0.014)
Observations	39,764	26,816	39,764	26,816
Specification	IV	IV	OLS	OLS
Sample	All	In U.S. >4 Years	All	In U.S. >4 Years

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Note: This table contains results from specifications that drop states that do not adjust their nominal benefit levels over time according to stated rules. Columns 1 and 2 use an instrumental variables specification, instrumenting for TANF maximum benefit levels with benefit levels calculated using stated rules. Columns 3 and 4 show results from Equation 1.1 with the same sample exclusions. The regressions include controls for marital status, sex, age and its square, number of own children in the household, unemployment rate in the year of arrival and the year of interview, and dummy variables for each of 14 sending countries, defined as in Table A2. Columns 2 and 4 exclude refugees who have been in the U.S. for fewer than five years, as they may still receive TANF benefits. All specifications control for year of arrival, year of interview, and state fixed effects, as well as state time trends. All standard errors are clustered at the state level.

Table A4: Mechanisms: Effects of TANF Generosity, Removing Controls

	Log Wage	
	(1)	(2)
TANF Max Benefit	0.069***	0.096***
	(0.022)	(0.026)
< High School	-0.560***	
	(0.035)	
High School	-0.441***	
	(0.029)	
Some College	-0.353***	
	(0.052)	
English Ability		0.324***
		(0.018)
Observations	34,479	34,479
Controls Excluded	English	Education

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Note: This table contains results from Equation 1.1 on the sample of refugees arriving since 1996, excluding those with 0 earnings. Column 1 excludes the control for English ability from the baseline specification, and column 2 excludes controls for educational attainment. The regressions include controls for marital status, age and its square, number of own children in the household, unemployment rate in the year of arrival and the year of interview, and dummy variables for each of 14 sending countries, defined as in Table A2. The specifications control for year of arrival, year of interview, and state fixed effects, as well as state time trends. All standard errors are clustered at the state level.

Table A5: Heterogeneity in Wage Effects of Cash Generosity

Dependent Variable:	Log Wage		
	(1)	(2)	(3)
TANF Max Benefit	0.062**	0.053*	0.085***
	(0.030)	(0.030)	(0.026)
< High School	-0.467***	-0.467***	-0.466***
	(0.033)	(0.033)	(0.033)
High School	-0.388***	-0.390***	-0.388***
	(0.031)	(0.032)	(0.032)
Some College	-0.332***	-0.333***	-0.332***
	(0.051)	(0.052)	(0.052)
English Ability	0.119**	0.220***	0.220***
	(0.053)	(0.014)	(0.014)
TANF*English	0.020**		
	(0.010)		
Age	0.055***	0.052***	0.055***
	(0.006)	(0.007)	(0.006)
TANF*Age		0.001**	
		0.000	
# of Children	-0.019**	-0.019**	0.009
	(0.009)	(0.009)	(0.018)
TANF*# of Children			-0.005*
			(0.003)
Observations	34,479	34,479	34,479
Interaction	English Ability	Age	# of Children

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Note: This table contains results from Equation 1.1 on the sample of refugees arriving since 1996, excluding those with 0 earnings. Each column shows the results with the generosity measure interacted with a refugee characteristic: column 1 shows interactions with English ability, column 2 shows an interaction with age, and column 3 shows an interaction with number of children. The regressions include controls for marital status, age and its square, number of own children in the household, unemployment rate in the year of arrival and the year of interview, and dummy variables for each of 14 sending countries, defined as in Table A2. The specifications control for year of arrival, year of interview, and state fixed effects, as well as state time trends. All standard errors are clustered at the state level.

Table A6: Effect of TANF Generosity on Wages, Adjusting for Regional Price Differences

Dependent Variable:	Log Wage			
	(1)	(2)	(3)	(4)
Adjusted TANF Max Benefit	0.037	0.07	0.059**	0.096***
	(0.037)	(0.046)	(0.022)	(0.032)
< High School			-0.392***	-0.467***
			(0.025)	(0.033)
High School			-0.334***	-0.389***
			(0.030)	(0.032)
Some College			-0.291***	-0.333***
			(0.050)	(0.052)
English Ability			0.233***	0.220***
			(0.015)	(0.014)
Observations	50,878	34,479	50,878	34,479
Controls	No	No	Yes	Yes
State Time Trends	All	In U.S. >4 Years	All	In U.S. >4 Years

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Note: This table contains results from Equation 1.1 on the sample of refugees arriving since 1996, excluding those with 0 earnings. In these specifications, I adjust my measure of TANF generosity using the BEA's Implicit Regional Price Deflator Series to reflect regional price differences. The regressions in columns 3 and 4 also include controls for marital status, sex, age and its square, number of own children in the household, unemployment rate in the year of arrival and the year of interview, and dummy variables for each of 14 sending countries. Columns 2 and 4 exclude refugees who have been in the U.S. for fewer than five years, as they may still receive TANF benefits. All specifications control for year of arrival, year of interview, and state fixed effects, as well as state time trends. All standard errors are clustered at the state level.

Table A7: Effects of TANF Maximum Benefit for a Family of Four on Wages

Dependent Variable:	Log Wage			
	(1)	(2)	(3)	(4)
TANF Max Benefit (Family of 4)	0.023	0.027	0.040***	0.053**
	(0.022)	(0.027)	(0.014)	(0.021)
< High School			-0.392***	-0.468***
			(0.025)	(0.033)
High School			-0.334***	-0.389***
			(0.030)	(0.032)
Some College			-0.291***	-0.333***
			(0.051)	(0.052)
English Ability			0.232***	0.219***
			(0.015)	(0.014)
Observations	50,733	34,369	50,733	34,369
Controls	No	No	Yes	Yes
Sample	All	In U.S. >4 Years	All	In U.S. >4 Years

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Note: This table contains results from Equation 1.1 on the sample of refugees arriving since 1996, using maximum benefit levels for a family of four rather than a family of three in the refugees' year of arrival. The regressions exclude refugees with 0 earnings. The regressions in columns 3 and 4 also include controls for marital status, sex, age and its square, number of own children in the household, unemployment rate in the year of arrival and the year of interview, and dummy variables for each of 14 sending countries. Columns 2 and 4 exclude refugees who have been in the U.S. for fewer than five years, as they may still receive TANF benefits. All specifications control for year of arrival, year of interview, and state fixed effects, as well as state time trends. All standard errors are clustered at the state level.

Table A8: Effects of TANF Generosity on Wages, Using Cutoff of 40 Percent

Dependent variable:	Log Wage			
	(4)	(4)	(4)	(4)
TANF Max Benefit	0.031	0.045**	0.017	0.036*
	(0.024)	(0.022)	(0.015)	(0.020)
< High School			-0.429***	-0.504***
			(0.033)	(0.041)
High School			-0.379***	-0.438***
			(0.041)	(0.043)
Some College			-0.335***	-0.379***
			(0.052)	(0.052)
English Ability			0.244***	0.231***
			(0.017)	(0.018)
Observations	67,331	47,579	67,331	47,579
Controls	No	No	Yes	Yes
Sample	All	In U.S. >4 Years	All	In U.S. >4 Years

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Note: This table contains results from Equation 1.1 on the sample of refugees arriving since 1996, excluding those with 0 earnings. These specifications use a cutoff of 40 percent for the definition of likely refugees, rather than 60 percent as in the rest of the analysis. The regressions in columns 3 and 4 also include controls for marital status, sex, age and its square, number of own children in the household, unemployment rate in the year of arrival and the year of interview, and dummy variables for each of 14 sending countries, defined as in Table A2. Columns 2 and 4 exclude refugees who have been in the U.S. for fewer than five years, as they may still receive TANF benefits. All specifications control for year of arrival, year of interview, and state fixed effects, as well as state time trends. All standard errors are clustered at the state level.

Table A9: Balance Table: Immigrant Characteristics Significantly Associated with TANF Generosity

	TANF Maximum Benefit
Dependent Variable	(1)
Marital Status	-0.004
	(0.003)
Female	-0.001
	(0.002)
Age	-0.025
	(0.079)
Number of Own Children	0.0189***
	(0.006)
< High School	0.0126***
	(0.003)
High School	0.000
	(0.002)
Some College	-0.00223**
	(0.001)
College and Above	-0.0103***
	(0.004)
English Ability	-0.0143***
	(0.004)
Mexico	0.0125***
	(0.004)
Canada	0.000
	(0.002)
Latin America	0.002
	(0.003)
N. & W. Europe	-0.001
	(0.001)
East Europe	-0.00902**
	(0.004)
East Asia	-0.00966**
	(0.004)
Southeast Asia	0.00644**
	(0.003)
Southwest Asia	0.000
	(0.002)
Middle East	-0.001
	(0.001)
Africa	0.001
	(0.001)
Oceania	-0.001
	(0.001)
Observations	1,411,003

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Note: This table contains results from Equation 1.1, run without controls, on the sample of non-refugee immigrants arriving since 1996. The outcome variables are immigrant characteristics. All specifications control for year of arrival, year of interview, and state fixed effects, as well as state time trends. All standard errors are clustered at the state level.

Table A10: Effects of TANF Generosity on Wages, Controlling for Years in U.S.

Dependent Variable:	Log Wage			
	(1)	(2)	(3)	(4)
TANF Max Benefit	0.03	0.047	0.054***	0.078***
	(0.028)	(0.035)	(0.018)	(0.025)
Years in the U.S.	0.037***	0.017***	0.034***	0.014***
	-0.001	-0.002	(0.001)	(0.002)
< High School			-0.397***	-0.468***
			(0.025)	(0.033)
High School			-0.337***	-0.390***
			(0.030)	(0.031)
Some College			-0.291***	-0.333***
			(0.050)	(0.052)
English Ability			0.235***	0.220***
			0	0
Observations	50,878	34,479	50,878	34,479
Controls	No	No	Yes	Yes
Sample	All	In U.S. >4 Years	All	In U.S. >4 Years

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Note: This table contains results from Equation 1.1 on the sample of refugees arriving since 1996, excluding those with 0 earnings. These specifications exclude controls for year of interview and instead control for a linear measure of years since arrival to the U.S. The regressions in columns 3 and 4 also include controls for years in the U.S., marital status, sex, age and its square, number of own children in the household, unemployment rate in the year of arrival and the year of interview, and dummy variables for each of 14 sending countries, defined as in Table A2. Columns 2 and 4 exclude refugees who have been in the U.S. for fewer than five years, as they may still receive TANF benefits. All specifications control for year of arrival and state fixed effects, as well as state time trends. All standard errors are clustered at the state level.

Table A11: Effects of TANF Generosity on Wages, Triple Difference Model

Dependent Variable:	Log Wage			
	(1)	(2)	(3)	(4)
TANF Max Benefit	-0.025***	-0.022***	-0.009**	-0.011***
	(0.008)	(0.007)	(0.004)	(0.003)
Refugee * TANF Max Benefit	0.047*	0.064*	0.059***	0.082***
	(0.023)	(0.036)	(0.016)	(0.030)
High School			0.095***	0.087***
			(0.008)	(0.008)
Refugee * High School			-0.008	-0.007
			(0.019)	(0.019)
Some College			0.189***	0.192***
			(0.012)	(0.012)
Refugee * Some College			-0.055**	-0.054**
			(0.025)	(0.025)
College			0.700***	0.733***
			(0.024)	(0.022)
Refugee * College			-0.268***	-0.265***
			(0.025)	(0.025)
English Ability			0.260***	0.246***
			(0.011)	(0.011)
Refugee * English Ability			-0.025***	-0.023**
			(0.009)	(0.009)
Observations	996,175	690,490	995,220	689,828
Individual Controls	No	No	Yes	Yes
Sample	All	In U.S. >4 Years	All	In U.S. >4 Years

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Note: This table contains results from a variant of Equation 1.1 using non-refugee immigrants as a control in a triple difference specification. The regression is run on the sample of immigrants arriving since 1996, excluding those with 0 earnings. The regressions in columns 3 and 4 also include controls for marital status, sex, age and its square, number of own children in the household, unemployment rate in the year of arrival and the year of interview, and dummy variables for each of 11 sending regions. These regions are: Mexico, Canada, Latin America (excluding Mexico), Northern and Western Europe, Eastern Europe, East Asia, Southeast Asia, Southwest Asia, the Middle East, Africa, and Oceania. All controls are fully interacted with refugee status. Columns 2 and 4 exclude immigrants who have been in the U.S. for fewer than five years. All specifications control for year of arrival and state fixed effects, as well as state time trends, all of which are interacted with refugee status. All standard errors are clustered at the state level.

Table A12: Heterogeneity in Effects of TANF Generosity on Non-Refugee Immigrants

Dependent Variable:	Log Wage		Employment	
	(1)	(2)	(3)	(4)
TANF Max Benefit	-0.023***	0.003	0.002	0.003**
	(0.006)	(0.007)	(0.002)	(0.001)
< High School	-0.732***	-0.570***	-0.105***	-0.101***
	(0.022)	(0.072)	(0.004)	(0.005)
High School	-0.645***	-0.533***	-0.071***	-0.063***
	(0.018)	(0.060)	(0.003)	(0.007)
Some College	-0.541***	-0.449***	-0.052***	-0.042***
	(0.019)	(0.059)	(0.004)	(0.010)
English Ability	0.132***	0.245***	0.057***	0.064***
	(0.047)	(0.010)	(0.008)	(0.004)
TANF*English	0.019***		0.001	
	(0.007)		(0.001)	
TANF*< High School		-0.027***		-0.001
		(0.010)		(0.001)
TANF* High School		-0.018**		-0.001
		(0.008)		(0.001)
TANF* Some College		-0.015*		-0.002
		(0.008)		(0.002)
Observations	656,011	656,011	938,872	938,872
Interaction	English Ability	Education	English Ability	Education

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Note: This table contains results from Equation 1.1 on the sample of non-refugee foreign born arriving since 1996, excluding those with 0 earnings in columns 1 and 2. The outcome variable in columns 1 and 2 is log wage, whereas in columns 3 and 4 a dummy for employment status is the outcome variable. Columns 1 and 3 also present an interaction between the measure of welfare generosity and English ability, while columns 2 and 4 show an interaction with educational attainment. The regressions also include controls for marital status, sex, age and its square, number of own children in the household, unemployment rate in the year of arrival and the year of interview, and dummy variables for each of 11 sending regions. These regions are Mexico, Canada, Latin America (excluding Mexico), Northern and Western Europe, Eastern Europe, East Asia, Southeast Asia, Southwest Asia, the Middle East, Africa, and Oceania. Columns 2 and 4 exclude refugees who have been in the U.S. for fewer than five years. All specifications control for year of arrival, year of interview, and state fixed effects, as well as state time trends. All standard errors are clustered at the state level.

Table A13: Effects of TANF Generosity on Low-Skilled Native Workers

Dependent Variable:	Log Wage		Employment	
	(1)	(2)	(3)	(4)
TANF Max Benefit	0.015	0.011	0.001	0
	(0.011)	(0.009)	(0.001)	(0.001)
< High School		-0.640***		-0.199***
		-0.015		(0.005)
Marital Status		0.202***		0.043***
		-0.005		(0.002)
Female		-0.486***		-0.078***
		-0.009		(0.002)
Age		0.113***		0.015***
		-0.001		0.000
Observations	19,638,249	19,638,249	32,394,690	32,394,690
Controls	No	Yes	No	Yes

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Note: This table contains results from Equation 1.1 on the sample of low-skilled natives, excluding those with 0 earnings in columns 1 and 2. The regressions show the effect of cash assistance in the year of interview on wages and the probability of employment. The regressions in columns 3 and 4 also include controls for marital status, sex, age and its square, number of own children in the household, and unemployment rate in the year of interview. All specifications control for year of interview fixed effects and state fixed effects, as well as state time trends (trend is in year of interview in this case). All standard errors are clustered at the state level.

Table A14: Effects of TANF Generosity Using Alternative Sample Definitions

Dependent Variable:	Log Wage	
	(1)	(2)
TANF Max Benefit	0.077***	0.084***
	(0.03)	(0.02)
< High School	-0.468***	-0.470***
	(0.03)	(0.04)
High School	-0.389***	-0.392***
	(0.03)	(0.03)
Some College	-0.333***	-0.335***
	(0.05)	(0.05)
English Ability	0.220***	0.222***
	(0.01)	(0.01)
Observations	34,479	35,419
Sample	Smoothed, Cutoff	Unsmoothed, Cutoff

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Note: This table contains results from Equation 1.1 on the sample of refugees arriving since 1996, excluding those with 0 earnings. Column 1 shows results using the baseline definition for refugee, while column 2 shows results where the fraction of refugees arriving in a year used for the sample cutoff is not smoothed. The regressions include controls for marital status, sex, age and its square, number of own children in the household, unemployment rate in the year of interview and year of arrival, and dummy variables for each of 14 sending countries, defined as in Table A2. All specifications control for year of interview fixed effects and state fixed effects, as well as state time trends. All standard errors are clustered at the state level.

Table A15: Test for Selection: TANF Maximum Benefit Levels Largely Do Not Predict Employed Refugee Characteristics

	TANF Maximum Benefit	
Dependent Variable	(1)	(2)
Marital Status	-0.012	0.007
	(0.011)	(0.021)
Female	0.010	0.032**
	(0.010)	(0.015)
Age	0.247	0.399
	(0.261)	(0.384)
Number of Own Children	-0.010	0.113**
	(0.053)	(0.045)
< High School	0.020**	0.018
	(0.009)	(0.012)
High School	-0.026	-0.041
	(0.033)	(0.046)
Some College	0.012	0.008
	(0.013)	(0.016)
College and Above	-0.005	0.016
	(0.030)	(0.038)
English Ability	-0.028***	-0.041
	(0.010)	(0.030)
Soviet Union	-0.089	-0.066
	(0.113)	(0.124)
Yugoslavia	-0.013	-0.016
	(0.069)	(0.100)
Southeast Asia	0.043	0.030*
	(0.017)	(0.016)
Middle East	-0.011	0.013
	(0.023)	(0.027)
Cuba	0.034	0.008
	(0.037)	(0.031)
Africa	0.036	0.031
	(0.043)	(0.042)
Observations	50,878	34,479
Sample	All	In U.S. >4 Years

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Note: This table contains results from Equation 1.1, run without controls, on the sample of employed refugees arriving since 1996. The outcome variables are refugee characteristics. In column 1, all employed refugees are included in the sample, while in column 2, I exclude refugees that arrived in the last 5 years before the interview. All specifications control for year of interview fixed effects and state fixed effects, as well as state time trends (trend is in year of interview in this case). All standard errors are clustered at the state level.

Table A16: Heckman Selection Model

Dependent variable:	Log Wage	Selection
	(1)	(2)
TANF Max Benefit	0.042**	0.005
	(0.015)	(0.027)
< High School	-0.399***	-0.572***
	(0.023)	(0.027)
High School	-0.336***	-0.284***
	(0.029)	(0.018)
Some College	-0.292***	-0.151***
	(0.050)	(0.017)
English Ability	0.234***	0.284***
	(0.015)	(0.029)
Observations	50,878	78,129
ρ	0.023*	
	(0.013)	
λ	0.02*	
	(0.011)	

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Note: This table contains results from a Heckman two step procedure, where I specify the availability of wage information to depend on cash assistance generosity, along with the usual set of controls. ρ gives an estimate of the correlation of the error terms of the wage and employment equations, while λ gives an estimate of ρ multiplied by the variance of the distribution of errors in the wage regression. All specifications control for year of arrival, year of interview, and state fixed effects, was well as state time trends. All standard errors are clustered at the state level.

A.0.1 Theoretical Appendix

This section presents a model of human capital investment for refugees that receive temporary cash assistance upon arrival to the U.S. For simplicity, and following Cortes (2004), suppose that the agents live for two periods after arrival in the U.S. Let utility be strictly increasing and concave in earnings (wage plus benefits). Assume that in the first period the refugees receive a cash benefit, whose level is determined exogenously. The refugees also spend a share θ of their time acquiring U.S.-specific human capital and the remainder of their time working in a job earning a wage w, with total wage earnings depending on initial human capital, H_0.[1] In the second period, the benefits expire, so that the refugees consume only what they earn in wages. However, the wage earnings in the second period are a function of the refugees' cumulative human capital, which depends on H_0 and the human capital they accumulated in the first period according to a strictly concave human capital production function f. The refugees maximize intertemporal utility in the two periods with a discount factor of $0 < \beta < 1$:

$$\max_{\theta} u(y_1) + \beta u(y_2) \tag{A.1}$$

where $y_1 = wH_0(1-\theta) + b$ and $y_2 = w(H_0 + f(H_0, \theta))$

The first order condition for maximization is:

$$-wH_0 u'(y_1) + \beta u'(y_2) \frac{\partial f(H_0, \theta)}{\partial \theta} w = 0 \tag{A.2}$$

[1]Intuitively, and similar to Cortes (2004), this represents the fact that human capital acquired abroad is not fully valued by U.S. employers, leading to underemployment among skilled refugees and other immigrants.

From which we can determine the effect of b on the optimal choice of investment θ:

$$\frac{\partial\theta^*}{\partial b} = \frac{wH_0u''(y_1)}{w^2{H_0}^2u''(y_1) + \beta u'(y_2)\frac{\partial^2 f(H_0,\theta)}{\partial\theta^2}w + \beta u''(y_2)(\frac{\partial f(H_0,\theta)}{\partial\theta})^2w^2} > 0 \tag{A.3}$$

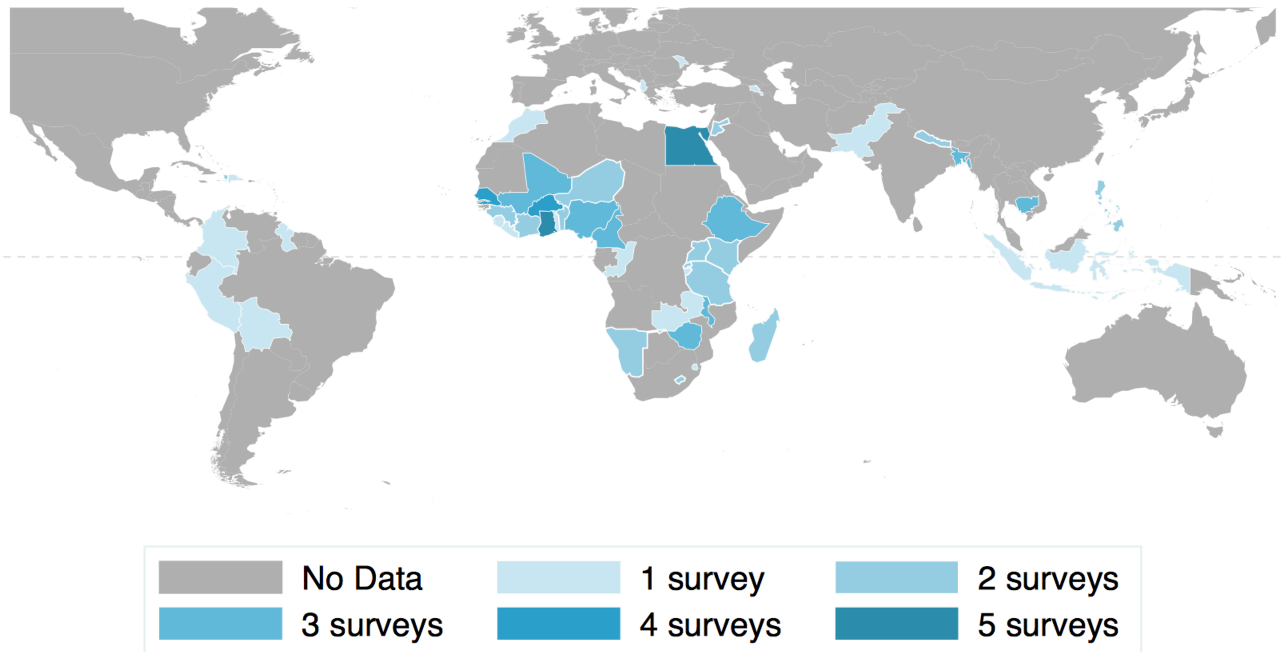
by the Implicit Function Theorem. The sign of the effect of b follows from the strict concavity of the utility and human capital production functions and from the fact that they are strictly increasing. Therefore, from this simplified model we obtain the result that an exogenous increase in cash benefits should result in an increase in human capital investment, resulting in increased wage earnings, an implication that can be tested empirically. This model also predicts by construction that refugees increase human capital investment by reducing labor supply, either on the intensive or extensive margin.

The unambiguous prediction of positive effects on wages relies on the fact that there is no impact of wages on receipt of benefits in this model. This is an abstraction from reality; earning high wages after resettlement results in reduced or terminated cash benefits. I argue that this assumption is reasonable in this context, since most refugees are no longer using cash assistance (excluding SSI) by three years after resettlement and human capital investment is a long-term consideration.

Appendix B

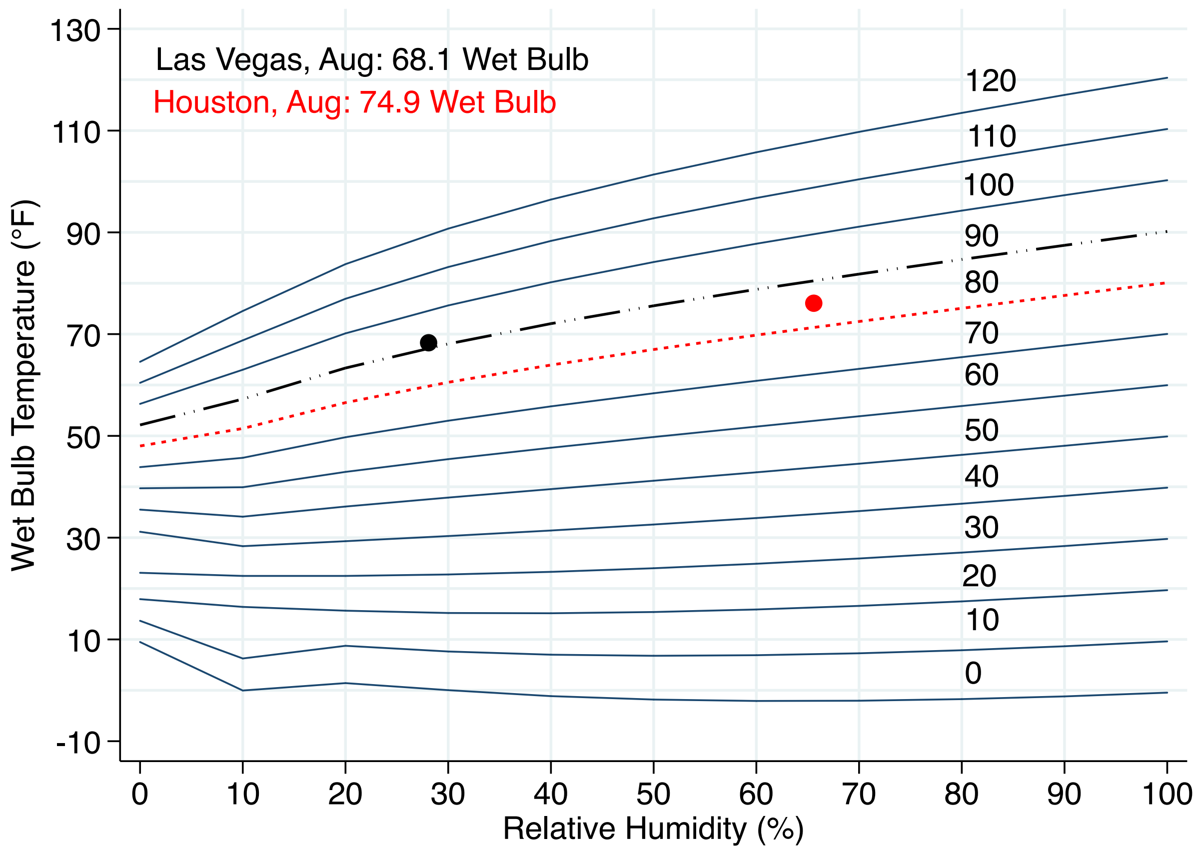
Appendix to Chapter 2

Figure B1: DHS Countries in Sample



Note: This figure shows the geographic location of countries where the DHS interviews were conducted. The shade of blue denotes the number of survey waves in the analysis.

Figure B2: Wet Bulb Temperature, Dry Bulb Temperature, and Humidity



Note: This figure displays the relationship between wet bulb temperature, dry bulb temperature, and relative humidity. In this figure, each line is an isometric line representing a fixed dry bulb temperature. The horizontal axis gives relative humidity, and the vertical axis gives wet bulb temperature. The two red points give August averages of daily average wet bulb temperatures in August, 2017 for Houston, TX and Las Vegas, NV, respectively, as illustrative examples.

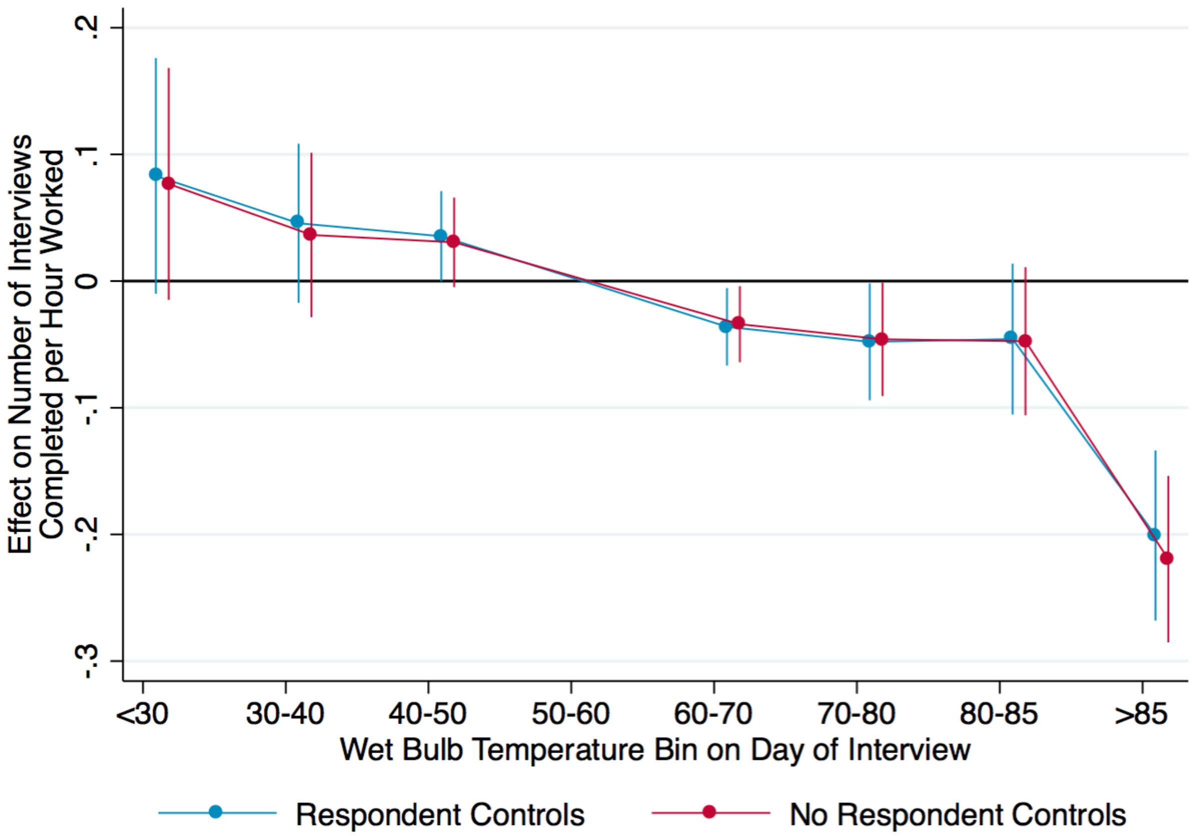
Figure B3: Questionnaire Example: Valid vs. Invalid Skips

SECTION 3. CONTRACEPTION

NO.	QUESTIONS AND FILTERS	CODING CATEGORIES	SKIP
302	CHECK 226: NOT PREGNANT OR UNSURE ☐ ↓	PREGNANT ☐	→ 312
303	Are you or your partner currently doing something or using any method to delay or avoid getting pregnant?	YES . . . . . 1 NO 2	 → 312
304 (4)	Which method are you using? RECORD ALL MENTIONED. IF MORE THAN ONE METHOD MENTIONED, FOLLOW SKIP INSTRUCTION FOR HIGHEST METHOD IN LIST.	FEMALE STERILIZATION . . . . . A MALE STERILIZATION . . . . . B IUD . . . . . C INJECTABLES . . . . . D IMPLANTS . . . . . E PILL . . . . . F CONDOM . . . . . G FEMALE CONDOM . . . . . H EMERGENCY CONTRACEPTION . . . . . I STANDARD DAYS METHOD . . . . . J LACTATIONAL AMENORRHEA METHOD . . . . . K RHYTHM METHOD . . . . . L WITHDRAWAL . . . . . M OTHER MODERN METHOD . . . . . X OTHER TRADITIONAL METHOD . . . . . Y	A–B → 307 C–E → 309 F G → 306 H–Y → 309
305	What is the brand name of the pills you are using? IF DON'T KNOW THE BRAND, ASK TO SEE THE PACKAGE.	BRAND A . . . . . 01 BRAND B . . . . . 02 BRAND C . . . . . 03 OTHER ________ 96 (SPECIFY) DON'T KNOW 98	→ 309

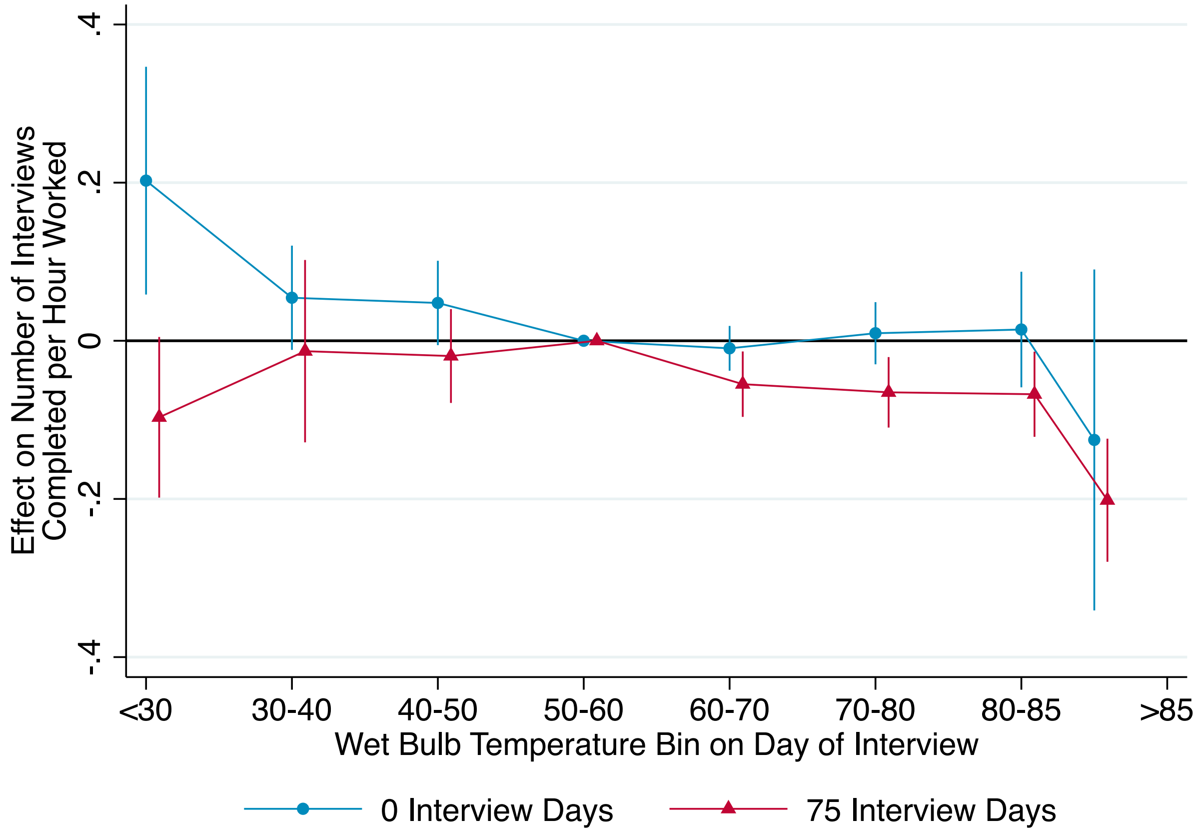
Note: This figure displays an illustrative example of the skip patterns in the DHS Wave 7 Individual Questionnaire. The section begins with instructions to check whether the respondent is pregnant from a previous question. If the respondent is pregnant, then questions 303-311 will be marked as valid skips. If the respondent is not pregnant, then the interviewer should move onto question 303. If the respondent is marked not pregnant or unsure but question 303 is left blank, then question 303 is a missing (or invalid missing) response.

Figure B4: Role of the Respondent: Respondent Controls Do Not Affect Main Results



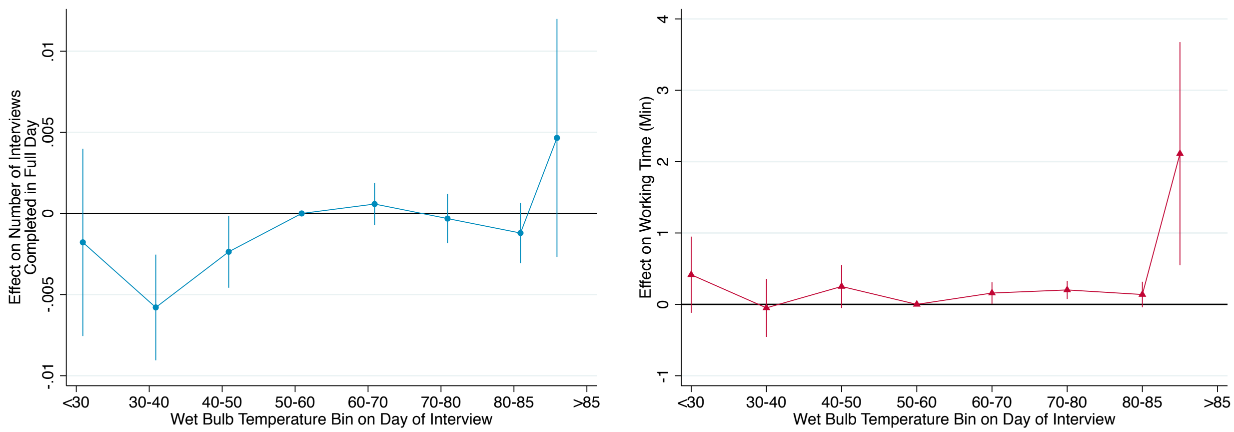
Note: This figure shows the results of interviewer-day level regressions using the number of interviews completed per hour worked as the outcome variable of interest, where hours worked is defined as the time between the start time of the first individual interview and the end time of the last individual interview in that day. The independent variables of interest are indicators for whether the daily average wet bulb temperature in the day of interview fell into the given bin. The blue line gives the full specification, including respondent control variables, while the red line omits respondent controls. The regressions also include fixed effects for the survey round by region of country as well as controls for the 10 year average of wet bulb temperature in the survey cluster in the month of interview, number of daylight hours, and the number of completed days in the survey round. Standard errors are clustered at the region-of-country level. Point estimates and 95% confidence intervals are shown.

Figure B5: Heterogeneity: The Effect of Heat on Number of Interviews Per Hour Increases with Experience (Interviewer FE)



Note: This figure shows the results of interviewer-day level regressions using the number of interviews completed per hour worked as the outcome variable of interest, where hours worked is defined as the time between the start time of the first individual interview and the end time of the last individual interview in that day. The independent variables of interest are indicators for whether the daily average wet bulb temperature in the day of interview fell into the given bin. Each wet bulb bin is interacted with a measure of how many days the interviewer has worked on that survey round. The blue line shows the effect for an interviewer on his/her first day, and the red line shows the effect for the interviewer's 75th day. The regressions also include interviewer fixed effects as well as controls for the characteristics of the set of respondents, the 10 year average of wet bulb temperature in the survey cluster in the month of interview, number of daylight hours, and the number of completed days in the survey round. Standard errors are clustered at the region-of-country level. Point estimates and 95% confidence intervals are shown.

Figure B6: More Experienced Interviewers Work Longer on Hot Days, Conduct Fewer Interviews on Cold Days (Interaction Effects)

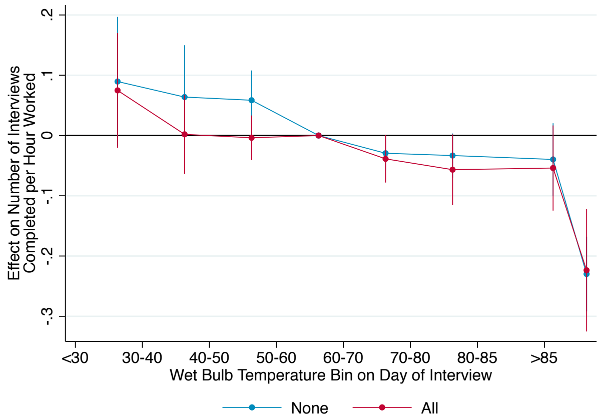


(a) Number of Interviews

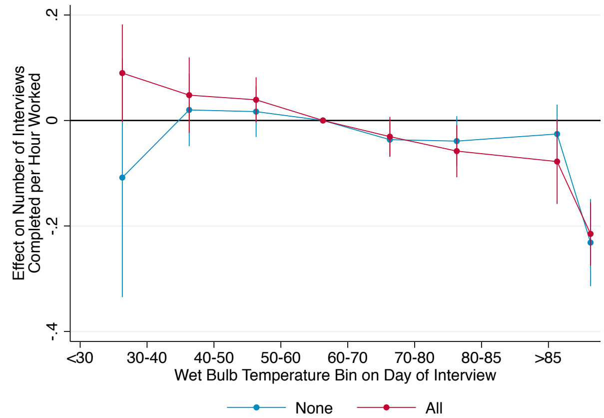
(b) Working Time

Note: This figure shows the results of interviewer-day level regressions using the number of interviews completed in a day (Panel A) and working time (Panel B) as the outcome variables of interest. The independent variables of interest are indicators for whether the daily average wet bulb temperature in the day of interview fell into the given bin. Each wet bulb bin is interacted with a measure of how many days the interviewer has worked on that survey round. The lines displayed show the estimated interaction effect: each coefficient is interpretable as the impact of a day of experience on the effect of wet bulb temperature on the outcome variable. The regressions also include region of country by survey round fixed effects as well as controls for the characteristics of the set of respondents, the 10 year average of wet bulb temperature in the survey cluster in the month of interview, number of daylight hours, and the number of completed days in the survey round. Standard errors are clustered at the region-of-country level. Point estimates and 95% confidence intervals are shown.

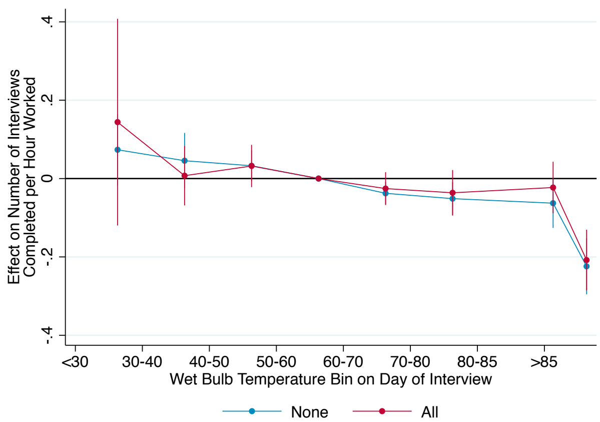
Figure B7: Role of the Respondent: Little Heterogeneity by Respondent Characteristics



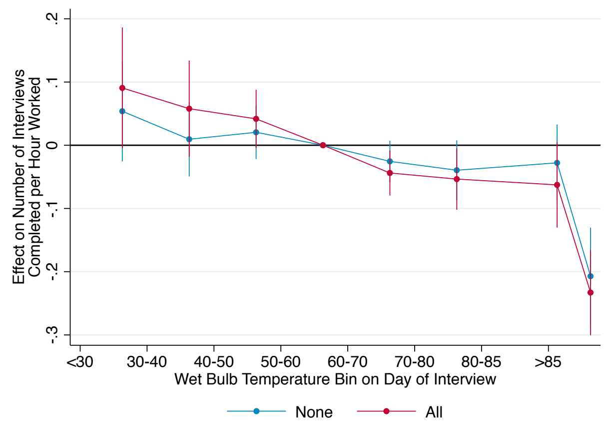
(a) Respondent Works



(b) HH Electricity



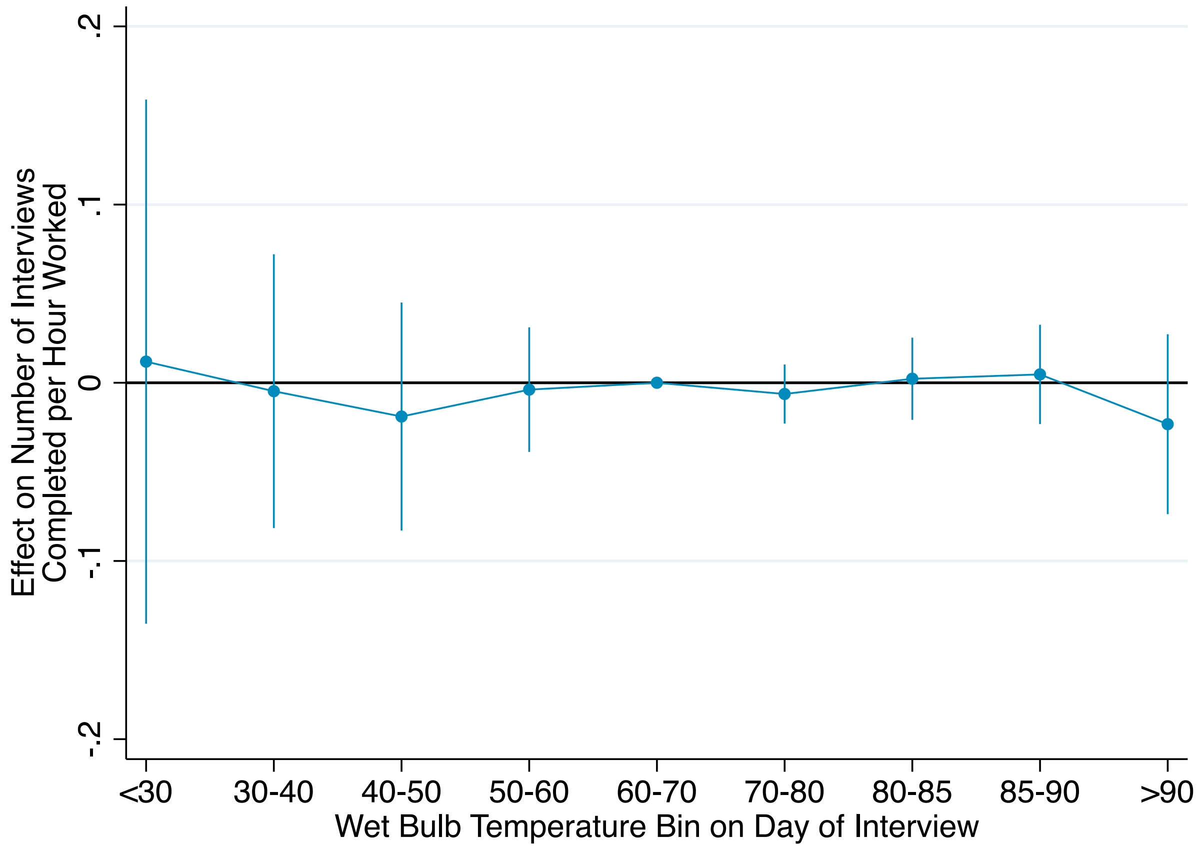
(c) Respondent Illiterate



(d) Formally Constructed House

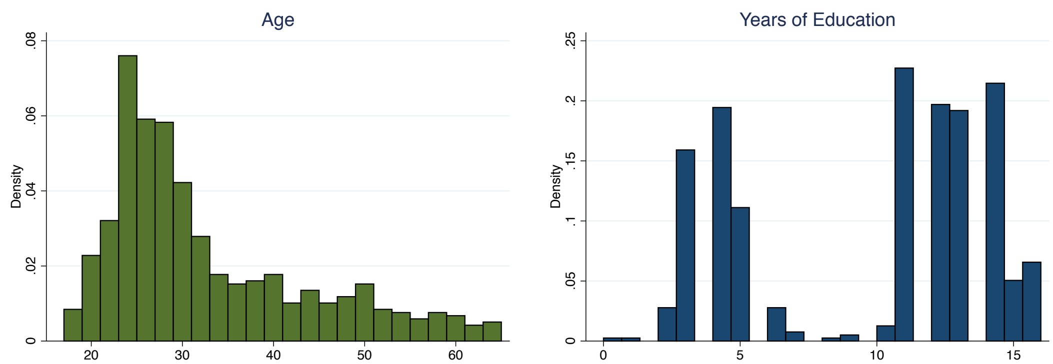
Note: This figure shows the results of interviewer-day level regressions using the number of interviews completed per hour worked as the outcome variable of interest, where hours worked is defined as the time between the start time of the first individual interview and the end time of the last individual interview in that day. Each panel shows interaction effects with the mix of respondents along one observable characteristic: these characteristics are whether the respondent works, whether the respondent's household has electricity, whether the respondent is illiterate, and whether the house is made of formal materials. The red line gives the effect of wet bulb temperature if all of the respondents in the interviewer-day have the characteristic, and the blue line gives the effect if none of the respondents have the characteristic. The regressions include fixed effects for the survey round by region of country as well as controls for other characteristics of the set of respondents, the 10 year average of wet bulb temperature in the survey cluster in the month of interview, number of daylight hours, and the number of completed days in the survey round. Standard errors are clustered at the region-of-country level. Point estimates and 95% confidence intervals are shown.

Figure B8: Alternative Specifications: Interviewer Productivity Does Not Respond to Intra-day Maximum Wet Bulb Temperature



Note: This figure shows the results of interviewer-day level regressions using the number of interviews completed per hour worked as the outcome variable of interest, where hours worked is defined as the time between the start time of the first individual interview and the end time of the last individual interview in that day. The independent variables of interest are indicators for whether the daily maximum wet bulb temperature in the day of interview fell into the given bin, where maximum temperature is defined as the highest of the eight daily readings in the Princeton Data. The regressions also include fixed effects for the survey round by region of country as well as controls for the characteristics of the set of respondents, the 10 year average of wet bulb temperature in the survey cluster in the month of interview, number of daylight hours, and the number of completed days in the survey round. Standard errors are clustered at the region-of-country level. Point estimates and 95% confidence intervals are shown.

Figure B9: Interviewer Characteristics in Post-2010 Data



Note: This figure contains summary statistics on interviewer characteristics from recent DHS surveys outside of the sample of the main analysis of this paper. These data come from fieldworker surveys included in the Afghanistan 2015-2016, Armenia 2015-2016, Nepal 2016 and Zimbabwe 2015 survey rounds. The left subfigure gives a histogram of the age distribution of fieldworkers, while the right subfigure gives the histogram of years of education.

Table B1: Interviewer Productivity Correlated with Probability of Leaving Survey Early

	Left Before Team					Days Between Last Day and Supervisor's Last Day				
Productivity Measure	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Invalid Missings	0.003***				0.003***	0.171**				0.195***
	-0.001				-0.001	-0.07				-0.07
Data Quality Flags		-0.010**			-0.006		-0.134			0.116
		-0.004			-0.004		-0.249			-0.247
Number of Interviews			-0.034***		-0.013***			-2.854***		-1.922***
			-0.003		-0.004			-0.153		-0.246
Working Time (Min)				-0.001***	-0.000***				-0.033***	-0.016***
				0	0				-0.002	-0.003
Observations	11,372	11,372	13,927	13,870	11,327	11,372	11,372	13,927	13,870	11,327

Clustered standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: This table shows the results of regressions showing the impact of average interviewer productivity on the probability of early separation from the survey team. Columns 1-5 examine the impact on the probability that the interviewer is last observed at least one day before the last member of their survey team. Columns 6-10 examine the impact on the number of days between the day that the interviewer is last observed and the day that the last member of their team is last observed. When data quality measures are included in the regressions, the number of questions asked in the interview is also included as a control, since longer interviews may mechanically contain more data quality issues. The regressions do not include other controls.

B.1 Data Appendix

This section details the construction of the weather dataset, as well as several key outcome variables in the data quality analysis.

B.1.1 Princeton Data

The weather data throughout the analysis come from the Princeton Meteorological Forcing Dataset, which is a reanalysis dataset that has been bias-corrected using observa-

Table B2: Wet Bulb Temperature Predicts Day of Survey Round

	Day of Round		
Wet Bulb Bin			
<30 Degrees	-5.940***	20.59***	45.61***
	(0.930)	(6.967)	(9.048)
30-40 Degrees	-4.913***	-0.403	16.94**
	(0.548)	(7.350)	(7.482)
40-50 Degrees	-26.26***	-17.85***	-7.034
	(0.248)	(5.066)	(4.290)
50-60 Degrees	-18.17***	-15.31***	-9.384***
	(0.157)	(3.036)	(2.535)
70-80 Degrees	9.939***	12.72***	7.174***
	(0.132)	(3.338)	(2.764)
80-85 Degrees	34.21***	12.21*	2.910
	(0.275)	(6.429)	(4.879)
>85 Degrees	79.08***	60.18***	42.96***
	(1.993)	(6.680)	(7.576)
Expected Temp			0.876**
Place FE		X	X
Observations	1,221,516	1,221,516	1,221,516

Note: This table shows the results of interview-level regressions using the day of the survey round (the number of days between the interview date and the date of the first interview in the survey round) as the outcome variable. The independent variables of interest are indicators for whether the daily average wet bulb temperature in the day of interview fell into the given bin. The first column just gives raw correlations, while the second column adds fixed effects for the survey round by region of country. The third column includes a control for the 10 year average of wet bulb temperature in the survey cluster in the month of interview Standard errors are clustered at the region-of-country level in columns 2 and 3.

Table B3: Mechanisms: More Data Quality Issues Arise on Hot Days

Dependent variable:	Quality Flags (1)	Valid Skips (2)	Invalid Missing (3)	Don't Know/Inconsistent
>85	0.282***	17.323	0.291	-0.303
	(0.084)	(15.616)	(0.233)	(0.211)
80-85	0.04	-2.259	0.436**	-0.077
	(0.037)	(3.298)	(0.191)	(0.135)
70-80	0.013	-4.575	0.347*	0.04
	(0.028)	(2.884)	(0.188)	(0.099)
60-70	-0.01	-2.59	0.292**	0.041
	(0.021)	(1.872)	(0.130)	(0.076)
40-50	0.100***	-1.225	0.051	-0.19
	(0.035)	(1.707)	(0.065)	(0.183)
30-40	0.081	1.529	0.02	-0.135
	(0.063)	(3.359)	(0.110)	(0.186)
<30	0.006	9.234**	-0.05	-0.089
	(0.078)	(3.959)	(0.266)	(0.238)
10-Yr Avg Wet Bulb	0.002	-0.469***	0.003	0
	(0.002)	(0.173)	(0.002)	(0.007)
Day of Survey Round	-0.000**	-0.066	-0.000**	0.001
	0.000	(0.045)	0.000	(0.001)
Daylight Hours	-0.001***	0.082***	-0.001***	0.002
	0.000	(0.023)	0.000	(0.001)
Region of Country FE	X	X	X	X
Observations	962,240	962,240	962,564	962,276

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Note: This table shows the results of interview-level regressions using counts of data quality problems as the outcome variables of interest. Column 1 examines total counts of data quality flags as the outcome variable, column 2 uses counts of valid skips (not applicable or not in universe), and column 3 uses counts of invalid missing responses. Column 4 uses a count of responses where the respondent did not know the answer to the question or where the response given was inconsistent with another response. More details on the construction of the outcome variables are available in Appendix B.1. The independent variables of interest are indicators for whether the daily average wet bulb temperature in the day of interview fell into the given bin. All regressions also include fixed effects for the survey round by region of country, controls for characteristics of the set of respondents, the 10-year average of wet bulb temperature in the survey cluster in the month of interview, number of daylight hours, and the number completed days in the survey round. Standard errors are clustered at the region-of-country level.

Table B4: Heterogeneity by World Region and Usual Climate

Dependent variable:	Number of Interviews Completed Per Hour Worked								
	Africa	East Asia Pacific	Europe & C. Asia	Latin America	Middle East & N. Africa	South Asia	Coolest Third	Middle Third	Warmest Third
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
>85	-	-	-	-0.129	-	-0.045	-	-	-0.196**
				(0.168)		(0.040)			(0.096)
80-85	-0.065*	-0.087**	-	0.023	-	0.006	-	0.125**	-0.095
	(0.038)	(0.039)		(0.103)		(0.025)		(0.063)	(0.088)
70-80	-0.059*	-0.044	-0.053	-0.019	-0.100*	0.036*	-0.206***	-0.059	-0.068
	(0.033)	(0.029)	(0.063)	(0.060)	(0.059)	(0.018)	(0.014)	(0.051)	(0.085)
60-70	-0.048***	-	-0.014	-0.034	-0.086**	0.038**	-0.022	-0.071	-
	(0.018)		(0.009)	(0.045)	(0.042)	(0.016)	(0.014)	(0.050)	
40-50	0.007	-	0.039	0.009	0.032	-0.056***	0.034*	-	-
	(0.030)		(0.051)	(0.035)	(0.035)	(0.011)	(0.020)		
30-40	0.008	-	0.08	0.009	-0.157*	-0.004	0.042	-	-
	(0.075)		(0.073)	(0.058)	(0.089)	(0.039)	(0.039)		
<30	-	-	0.124	0.004	-0.192***	-0.071**	0.103*	-	-
			(0.101)	(0.077)	(0.064)	(0.031)	(0.057)		
10-Yr Avg Wet Bulb	0.004*	0.005	0.015	-0.001	-0.005	-0.003	0.001	0.004	0.007**
	(0.002)	(0.004)	(0.012)	(0.003)	(0.005)	(0.002)	(0.002)	(0.003)	(0.003)
Day of Survey Round	0.001**	0.000**	0.003**	0.001***	0.007***	0.001*	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.002)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Daylight Hours	0	0.001***	-0.001	0.001***	0.001	0	0	-0.001**	0
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)
Observations	175,394	39,511	7,720	43,428	21,770	26,906	106,861	166,953	94,955

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Note: The first 6 columns of this table show the results of interviewer-day level regressions using the number of interviews completed per hour worked as the outcome variable of interest, run separately for six world regions. Hours worked is defined as the time between the start time of the first individual interview and the end time of the last individual interview in that day. The independent variables of interest are indicators for whether the daily average wet bulb temperature in the day of interview fell into the given bin. The regressions also include fixed effects for the survey round by region of country as well as controls for the characteristics of the set of respondents, the 10 year average of wet bulb temperature in the survey cluster in the month of interview, number of daylight hours, and the number of completed days in the survey round. Standard errors are clustered at the region-of-country level. Table B6 shows the list of countries in each region. Columns 7 through 9 show the results of similar subsample regressions: these regressions break the sample into three groups based on annual wet bulb temperature.

Table B5: Daylight Hours Do Not Mediate Relationship Between Temperature and Start Time

	Start Time		
Wet Bulb Bin			
<30 Degrees		105.489***	111.552***
		(4.193)	(4.298)
30-40 Degrees		59.637***	63.555***
		(2.489)	(2.563)
40-50 Degrees		17.653***	19.935***
		(1.252)	(1.301)
50-60 Degrees		-4.091***	-2.649***
		(0.769)	(0.801)
70-80 Degrees		-14.239***	-14.421***
		(0.664)	(0.664)
80-85 Degrees		-52.843***	-53.644***
		(1.407)	(1.412)
>85 Degrees		-163.049***	-166.374***
		(10.657)	(10.669)
Daylight	-0.099***		0.042
	(0.006)		(0.006)
Observations	413,417	413,417	413,417

Note: This table shows the results of interviewer-day level regressions using the starting time of the first interview of the day as the outcome variable. The independent variables of interest are indicators for whether the daily average wet bulb temperature in the day of interview fell into the given bin. The first column shows the correlation between daylight hours and starting time, while the second column gives the correlation between wet bulb temperature bins and starting time. The third column includes both temperature and daylight hours, showing that the inclusion of daylight hours into the regression does not significantly change the coefficients on the wet bulb bins.

tional data. Reanalysis datasets combine observational data from multiple sources (such as satellites, weather balloons, ground stations, etc.) with physics-based weather models that extend the data to observationally sparse geographies. The Princeton Data incorporates reanalysis data from the National Centers for Environmental Prediction and National Center for Atmospheric Research (the NCEP-NCAR Reanalysis Data) with observational data from the Climactic Research Unit (CRU) and the Global Precipitation Climatology Project (GPCP). For more details on the Princeton Data, see Justin Sheffield, Gopi Goteti and Eric F. Wood (2006). I, like Geruso and Spears (2018*a*), use data on dry bulb temperature, specific humidity, and pressure for 1990-2010 and calculate relative humidity and wet bulb temperature using the following calculations:

1. Relative humidity is calculated as follows, combining standard formulas for the mixing ratio, saturation mixing ratio, and specific humidity from the World Meteorological Organization:

$$rh = 0.263 * p * sh * \left[exp\left(\frac{17.67(t - 273.16)}{t - 29.65} \right) \right]^{-1} \tag{B.1}$$

   where rh is relative humidity (%), p is pressure (Pa), sh is specific humidity, and t is temperature (K).

2. Then, wet bulb temperature is calculated using the Stull Calculation, which is standard for sea-level pressure and uses temperature in degrees Celsius rather than Kelvin:

$$\begin{aligned} wb = t * [atan(0.151977 * (rh + 8.313658)^{\frac{1}{2}}] + atan(t + rh) - atan(rh - 1.676331) + \\ 0.00391838(rh)^{\frac{3}{2}} * atan(0.023101rh) - 4.686035 \end{aligned} \tag{B.2}$$

Once wet bulb temperature is calculated, the weather variables are merged with the DHS variables in the manner described in the text: the four surrounding grid points for each DHS survey cluster are located and merged with the weather data, and then the weather variables are calculated as the averages of the four surrounding grid points, weighted by inverse distance between each grid point and the survey cluster.

B.1.2 Counts of Data Quality Flags

There are three major types of data quality flags used in the analysis, as follows:

1. Imputed Dates: these are important dates where the information has been imputed, either because the full date was not recorded or because the date given was inconsistent with another date (for example, births that are less than 7 months apart).

    - Date of birth of the respondent
    - Date of first union or marriage
    - Date of birth of each child of the respondent
    - Date of conception of the current pregnancy
    - Date of start of use of current method of contraception
    - Date of last terminated pregnancy

2. Flagged body measurements: in the DHS, several body measurements are recorded. When they are out of the “acceptable” range of that measurement, they are flagged as such.

    - Child height or weight measurement

- Women's weight or height

3. Duration variables: the DHS asks for the duration of key reproduction-related activities, such as breastfeeding, and it includes flag variables for inconsistencies in the data processing phase.

    - Duration of breastfeeding
    - Duration of postpartum amenorrhea
    - Duration of postpartum abstinence
    - Time since last menstrual period
    - Time between first marriage and first birth (on a base of ever-married women)

The outcome variable for count of data quality flags gives the total quantity of these variables that are flagged. Since many respondents have more than one child, such that there is the potential for flags on important dates or measurements for multiple children, this variable ranges in practice from 0 to 23 for women. Most of these flags apply only to interviews with women (this applies to most duration and measurement variables). This variable for men therefore ranges from 0 to 2.

B.1.3 Missing Data

Missing data in the DHS is coded consistently across variables, and variable names are consistent within modules of the survey, allowing me to construct a count of total missing variables in each interview. There are four categories of codes, which are described below. The descriptions include the relevant information from the DHS Recode Manual.

1. Missing data: "This question should have been answered by the respondent, but the questionnaire contained no information for this variable." Depending on the range of codes for the variable, this is coded as 9, 99, 999, or 9999.

2. Respondent did not know the answer: "The respondent replied 'Don't know' to this question. Depending on the range of codes for the variable, this is coded as 8, 98, 998, or 9998. There are exceptions to this rule, which I accommodate to the best of my ability. These include the following:

   - In the contraception module, the questionnaire asks about knowledge and use of a range of methods of contraception. Some of these are country specific. If a method is not mentioned in a certain country's survey, it is coded as "8." I remove these from the count of "I don't know" answers and add them to the valid skips.
   - For several body measurement variables, measurements flagged as outside the usual range are coded as "9998" or "99998." I remove these from the count of "I don't know" answers and include them as data quality flags.

3. Inconsistent answer: "The answer to this question was inconsistent with other responses in the questionnaire and it was thought that this response was probably in error. The response was changed to this code to avoid further problems due to inconsistency of information. This usually takes place during the secondary editing stage of data processing." Depending on the range of codes for the variable, this is coded as 7, 97, 997, 9997. There are exceptions to this rule, including:

   - In the module on basic respondent data, the code 7 often means that the respondent is not a de jure resident of the household. In these cases, I remove the response from the count of inconsistent responses and add it to the count of valid skips.

4. Valid skips: "Variable is not applicable for this respondent either because the question was not asked in a particular country or because the question was not asked of the respondent due to the flow or skip pattern of the questionnaire." This is coded to be missing.

Most survey modules in the DHS are either used by every survey implementing country or are optional ones implemented by a substantial fraction of them. However, many countries also include a multitude of country-specific variables with a variety of names. The count of missing data instances in this paper encompass only standard variables, as defined by the names. The included modules are listed below with the relevant variable names. The variable names are usually built in 2 to 3 components. For three-component variables, the first component indicates the type of interview (v for women, hv or hc for household, mv for men), the second component indicates the module, and the third gives the variable number. For two-component variables, the first component indicates the module and the second the variable number. The two-component variables usually are for questions about children or special modules.

1. Respondent's basic data (v0's and v1's)

2. Reproduction (v2's and b's)

3. Contraceptive use (v3's)

4. Maternity (m1-m73)

5. Maternity and feeding (v4's)
6. Health history (of children under five) (h1-h22)
7. Height and weight (of children under five) (hw's)
8. Marriage (v5's)
9. Fertility preference (v6's)
10. Partner's characteristics and women's work (v71's through v74's)
11. AIDS, STI's, and condom use (v75's through v77's; v82's through v85's)
12. Interview characteristics (v80's and v81's)
13. Maternal mortality (mm's)
14. Malaria (ml's)
15. Domestic violence (d's)
16. Female genital cutting (g's)

B.1.3.1 Men's Interviews

1. Respondent's basic data (mv0's and mv1's)
2. Reproduction (mv2's)
3. Contraceptive use (mv3's)
4. Tuberculosis and other health issues (mv4's)

5. Marriage (mv5's)

6. Fertility preferences (mv6's)

7. Occupation and work status (mv71's through mv74's)

8. AIDS and condom use (mv75's through mv77's; mv82's through mv85's)

9. Female genital cutting (g's)

Table B6: DHS Survey Rounds in Sample

Country Name	Dates of Fieldwork	Number of Interviewers	Number of Regions	Number of Clusters	Region
Albania	10/2008-4/2009	61	4	450	Europe & Central Asia
Armenia	10/2010-12/2010	73	11	289	Europe & Central Asia
Bangladesh	11/1999-4/2000	95	6	341	South Asia
Bangladesh	1/2004-5/2004	115	6	359	South Asia
Bangladesh	3/2007-8/2007	106	6	361	South Asia
Burkina Faso	12/1992-4/1993	85	10	459	Africa
Burkina Faso	11/1998-3/1999	50	5	208	Africa
Burkina Faso	6/2003-12/2003	79	14	397	Africa
Burkina Faso	5/2010-12/2010	113	13	541	Africa
Benin	6/1996-8/1996	54	6	81	Africa
Benin	8/2001-11/2001	65	6	247	Africa
Bolivia	2/2008-6/2008	261	9	996	Latin America & Caribbean
Burundi	8/2010-12/2010	72	5	338	Africa
Central African Republic	9/1994-3/1995	43	6	224	Africa
Cote d'Ivoire	6/1994-11/1994	48	10	246	Africa
Cote d'Ivoire	9/1998-3/1999	41	3	140	Africa
Cameroon	4/1991-10/1991	72	10	258	Africa
Cameroon	2/2004-9/2004	90	12	464	Africa
Colombia	11/2009-12/2010	91	6	4846	Latin America & Caribbean
Dominican Republic	3/2007-8/2007	164	32	1425	Latin America & Caribbean
Egypt	11/1992-2/1993	167	5	546	Middle East & North Africa
Egypt	11/1995-2/1996	69	6	933	Middle East & North Africa
Egypt	2/2000-5/2000	144	6	998	Middle East & North Africa
Egypt	3/2008-6/2008	57	6	1243	Middle East & North Africa
Egypt	4/2005-7/2005	153	6	1298	Middle East & North Africa
Ethiopia	2/2000-6/2000	230	11	535	Africa
Ethiopia	4/2005-9/2005	178	11	528	Africa
Ethiopia	12/2010-12/2010	186	11	31	Africa
Ghana	10/1993-2/1994	109	20	533	Africa
Ghana	7/2003-11/2003	74	10	410	Africa
Ghana	9/2008-12/2008	111	10	404	Africa
Ghana	11/1998-2/1999	69	10	400	Africa
Guinea	4/1999-8/1999	54	5	293	Africa
Guinea	2/2005-6/2005	51	8	291	Africa
Guyana	3/2009-8/2009	99	10	312	Latin America & Caribbean
Haiti	2/2000-7/2000	65	10	316	Latin America & Caribbean
Haiti	10/2005-5/2006	65	10	332	Latin America & Caribbean
Indonesia	10/2002-4/2003	375	26	1317	East Asia Pacific
Jordan	7/2002-10/2002	70	3	495	Middle East & North Africa
Jordan	6/2007-10/2007	115	3	924	Middle East & North Africa
Kenya	4/2003-9/2003	98	8	399	Africa
Kenya	11/2008-3/2009	123	8	397	Africa

Cambodia	1/2000-7/2000	94	23	470	East Asia Pacific
Cambodia	9/2005-3/2006	92	19	548	East Asia Pacific
Cambodia	7/2010-12/2010	79	19	591	East Asia Pacific
Liberia	12/2006-4/2007	99	6	291	Africa
Lesotho	9/2004-2/2005	58	10	381	Africa
Lesotho	10/2009-2/2010	78	10	395	Africa
Morocco	10/2003-2/2004	58	15	479	Middle East & North Africa
Moldova	6/2005-8/2005	76	4	399	Europe & Central Asia
Madagascar	8/1997-12/1997	46	6	268	Africa
Madagascar	11/2008-7/2009	123	22	585	Africa
Mali	11/1995-5/1996	60	8	300	Africa
Mali	1/2001-6/2001	147	9	399	Africa
Mali	3/2006-12/2006	198	9	405	Africa
Malawi	7/2000-11/2000	168	3	559	Africa
Malawi	1/2004-2/2005	130	3	520	Africa
Malawi	6/2010-10/2010	302	3	827	Africa
Nigeria	4/1990-12/1990	170	4	297	Africa
Nigeria	3/2003-8/2003	77	6	360	Africa
Nigeria	6/2008-11/2008	254	6	886	Africa
Niger	3/1992-6/1992	52	8	235	Africa
Niger	1/1998-7/1998	64	6	268	Africa
Namibia	9/2000-12/2000	109	13	260	Africa
Namibia	11/2006-4/2007	194	13	490	Africa
Nepal	1/2001-7/2001	74	5	251	South Asia
Nepal	2/2006-8/2006	87	5	260	South Asia
Peru	7/2000-11/2000	211	24	1408	Latin America & Caribbean
Philippines	6/2003-9/2003	303	17	816	East Asia Pacific
Philippines	8/2008-9/2008	277	17	789	East Asia Pacific
Pakistan	9/2006-3/2007	130	4	954	South Asia
Rwanda	2/2005-8/2005	89	12	456	Africa
Rwanda	9/2010-12/2010	112	5	279	Africa
Sierra Leone	4/2008-9/2008	133	4	350	Africa
Senegal	11/1992-8/1993	51	8	508	Africa
Senegal	1/1997-5/1997	130	8	638	Africa
Senegal	1/2005-6/2005	77	11	366	Africa
Senegal	10/2010-12/2010	64	11	161	Africa
Swaziland	7/2006-3/2007	80	4	269	Africa
Togo	2/1998-5/1998	64	6	287	Africa
Timor-Leste	8/2009-2/2010	149	13	454	East Asia Pacific
Tanzania	9/1999-12/1999	78	22	173	Africa
Tanzania	12/2009-5/2010	70	26	458	Africa
Uganda	9/2000-3/2001	72	4	267	Africa
Uganda	5/2006-10/2006	87	9	336	Africa
Zambia	4/2007-10/2007	105	9	319	Africa
Zimbabwe	8/1999-12/1999	83	10	221	Africa
Zimbabwe	8/2005-4/2006	135	10	396	Africa
Zimbabwe	9/2010-12/2010	72	10	219	Africa

Note: This table gives descriptions of each of the 92 survey rounds used in the main analysis of the paper. Each row of the table represents one survey round; some countries have multiple. The second column gives the start and end date of the interviews. Column 3 gives the number of unique interviewers observed in the survey round, column 4 gives the number of regions of that country used in the round (used as fixed effects in the main analysis), and column 5 gives the number of clusters where interviews were conducted.

B.2 Theoretical Framework

In this section I present a simple framework for the interviewer's endogenous choice of effort allocation on days with varying temperatures. I closely follow the model set forth by Joshua Graff Zivin and Matthew Neidell (2012) to describe worker responses to air pollution in assuming that output for an interviewer is a function of effort e and temperature Ω. Here, a large value of Ω can be thought of a more extreme temperature. The interviewer's output has two components: data quality (q) and data quantity (y), and the interviewer chooses her effort allocation for both components (e_q and e_y, respectively). The interviewer earns a fixed wage $\bar{w}$ for as long as she works on the survey round and has a continuation value of K associated with keeping her job (not being fired for poor performance). The worker's probability of job retention depends linearly on output, for simplicity. The probability of job retention depends on both the quality and quantity of the data produced; each dimension is weighted by α and $1 - \alpha$ in the job retention decision, respectively.

Interviewers choose effort levels e_y and e_q in order to maximize the following:

$$\max_{e_y, e_q} \bar{w} + (\alpha y(e_y) + (1 - \alpha) q(e_q))K - c(e_y, e_q, \Omega) \tag{B.3}$$

The first order conditions are the following:

$$\alpha \frac{\partial y}{\partial e_y} K = \frac{\partial c}{\partial e_y} \tag{B.4}$$

$$(1 - \alpha) \frac{\partial q}{\partial e_q} K = \frac{\partial c}{\partial e_q} \tag{B.5}$$

Taking a total derivative with respect to Ω yields:

$$\alpha K(\frac{\partial^2 y}{\partial {e_y}^2})\frac{\partial e_y}{\partial \Omega} = \frac{\partial^2 c}{\partial {e_y}^2}\frac{\partial e_y}{\partial \Omega} + \frac{\partial c}{\partial e_y \partial \Omega} \tag{B.6}$$

$$(1-\alpha)K(\frac{\partial^2 q}{\partial {e_q}^2})\frac{\partial e_q}{\partial \Omega} = \frac{\partial^c}{\partial {e_q}^2}\frac{\partial e_q}{\partial \Omega} + \frac{\partial c}{\partial e_q \partial \Omega} \tag{B.7}$$

Solving for $\frac{\partial e_q}{\partial \Omega}$ and $\frac{\partial e_y}{\partial \Omega}$, respectively, yields the following relationships:

$$\frac{\partial e_y}{\partial \Omega} = \frac{\frac{\partial c}{\partial e_y \partial \Omega}}{\alpha K \frac{\partial^2 y}{\partial {e_y}^2} - \frac{\partial^2 c}{\partial {e_y}^2}} \tag{B.8}$$

$$\frac{\partial e_q}{\partial \Omega} = \frac{\frac{\partial c}{\partial e_q \partial \Omega}}{(1-\alpha) K \frac{\partial^2 q}{\partial {e_q}^2} - \frac{\partial^2 c}{\partial {e_q}^2}} \tag{B.9}$$

Assuming that the cost of effort function is convex, production is concave in effort, and that extreme temperature increases the marginal cost of effort, the effect of temperature on effort allocation is negative for both tasks. This model also predicts that the effect of temperature on effort allocation is decreasing in the weight put on the dimension of productivity in the probability of job retention. Similarly, it is decreasing in the continuation value of the job, K. The model therefore predicts that the negative effect of temperature on effort will be stronger on the less observable dimension of productivity, quality.

B.2.1 Alternative Assumptions

In the previous section, the only way in which temperature (Ω) entered the interviewer's decision was through disutility of effort. It may also be reasonable to assume that temperature also has a direct effect on productivity, through some unavoidable physiological effect. The maximization problem could be changed to reflect this possibility in the following way:

$$\max_{e_y,e_q} \bar{w} + (\alpha y(e_y,\Omega) + (1-\alpha)q(e_q,\Omega)K - c(e_y,e_q,\Omega) \tag{B.10}$$

The first order conditions remain the same, but taking the derivative with respect to temperature now yields:

$$\alpha K(\frac{\partial^2 y}{\partial {e_y}^2})\frac{\partial e_y}{\partial \Omega} + \alpha\frac{\partial y}{\partial e_y \partial \Omega}K = \frac{\partial^2 c}{\partial {e_y}^2}\frac{\partial e_y}{\partial \Omega} + \frac{\partial c}{\partial e_y \partial \Omega} \tag{B.11}$$

$$(1-\alpha)K(\frac{\partial^2 q}{\partial {e_q}^2})\frac{\partial e_q}{\partial \Omega} + \alpha\frac{\partial q}{\partial e_q \partial \Omega}K = \frac{\partial^c}{\partial {e_q}^2}\frac{\partial e_q}{\partial \Omega} + \frac{\partial c}{\partial e_q \partial \Omega} \tag{B.12}$$

The resulting expression for the effect of temperature on effort allocation is as follows:

$$\frac{\partial e_y}{\partial \Omega} = \frac{\frac{\partial c}{\partial e_y \partial \Omega} - \alpha K \frac{\partial y}{\partial e_y \partial \Omega}}{\alpha K \frac{\partial^2 y}{\partial {e_y}^2} - \frac{\partial^2 c}{\partial {e_y}^2}} \tag{B.13}$$

$$\frac{\partial e_q}{\partial \Omega} = \frac{\frac{\partial c}{\partial e_q \partial \Omega} - (1-\alpha) K \frac{\partial q}{\partial e_q \partial \Omega}}{(1-\alpha) K \frac{\partial^2 q}{\partial {e_q}^2} - \frac{\partial^2 c}{\partial {e_q}^2}} \tag{B.14}$$

The overall effects of temperature on effort allocation on both tasks remains negative, as long as temperature is assumed to decrease the marginal productivity of effort. The weight put on each task in the job retention probability function now appears both in the numerator and the denominator, however. The effect of an increase in the weight is to increase the size of both the numerator and the denominator in absolute value. Intuitively, temperature now decreases the returns to effort, and if this effect is large enough the differential effect on separate tasks may be swamped. Therefore, it is ultimately an empirical question whether temperature will have larger effects on the less observable dimension.

Bibliography

Adhvaryu, Achyuta, Namrata Kala, and Anant Nyshadham. 2018. "The Light and the Heat: Productivity Co-benefits of Energy-saving Technology." NBER Working Paper 24314.

Archsmith, James, Anthony Heyes, and Soodeh Saberian. 2018. "Air Quality and Error Quantity: Pollution and Performance in a High-Skilled, Quality-Focused Occupation." *Journal of the Association of Environmental and Resource Economists*, 5(4): 827–863.

Aslund, Olof, Per-Anders Edin, Peter Fredriksson, and Hans Grönqvist. 2011. "Peers, neighborhoods, and immigrant student achievement: evidence from a placement policy." *American Economic Journal: Applied Economics*, 3(2): 67–95.

Barreca, Alan. 2012. "Climate Change, Humidity, and Mortality in the United States." *Journal of Environmental Economics and Management*, 63(1): 19–34.

Beaman, Lori. 2012. "Social networks and the dynamics of labour market outcomes: evidence from refugees resettled in the U.S." *Review of Economic Studies*, 79(1): 128–161.

Behrer, Patrick, and Jisung Park. 2017. "Will We Adapt? Temperature, Labor and Adaptation to Climate Change." Working Paper.

Blank, Rebecca. 2002. "Evaluating welfare reform in the United States." *Journal of Economic Literature*, 40(4): 1105–1166.

Borjas, George J. 2000. "Ethnic enclaves and assimilation." *Swedish Economic Policy Review*, 7: 89–122.

Bruno, Andorra. 2011. "U.S. refugee resettlement assistance." Congressional Research Service, Washington, D.C.

Burke, Marshall, and Kyle Emerick. 2016. "Adaptation to Climate Change: Evidence from US Agriculture." *American Economic Journal: Economic Policy*, 8(3): 106–40.

Burke, Marshall, Solomon M. Hsiang, and Edward Miguel. 2015. "Global non-linear effect of temperature on economic production." *Nature*, 527: 235–239.

Capps, Randy, and Kathleen Newland. 2015. "The integration outcomes of U.S. refugees: successes and challenges." Migration Policy Institute, Washington, D.C.

Case, Anne, and Christina Paxson. 2008. "Stature and status: Height, ability, and labor market outcomes." *Journal of political Economy*, 116(3): 499–532.

Case, Anne, and Christina Paxson. 2011. "The long reach of childhood health and circumstance: evidence from the Whitehall II Study." *The Economic Journal*, 121(554): F183–F204.

Case, Anne, and Ta-Nehisi Coates. 2017. "Fear and despair: Consequences of inequity." In *Knowledge to Action: Accelerating Progress in Health, Well-Being, and Equity.* , ed. Alonzo L. Plough. Oxford:Oxford University Press.

Chen, Xiaoguang, and Lu Yang. 2017. "Temperature and industrial output: Firm-level evidence from China." *Journal of Environmental Economics and Management.*

Chetty, Raj. 2008. "Moral hazard vs. liquidity and optimal unemployment insurance." *Journal of Political Economy*, 116(2): 173–234.

Chinkhumba, Jobiba, Susan Godlonton, and Rebecca Thornton. 2014. "The Demand for Medical Male Circumcision." *American Economic Journal: Applied Economics*, 6(2): 152–177.

Coffey, Diane. 2015. "Prepregnancy body mass and weight gain during pregnancy in India and sub-Saharan Africa." *Proceedings of the National Academy of Sciences*, 112(11): 3302–3307.

Coffey, Diane, Aashish Gupta, Payal Hathi, Nidhi Khurana, Dean Spears, Nikhil Srivastav, and Sangita Vyas. 2014. "Revealed preference for open defecation." *Economic & Political Weekly*, 49(38): 43.

Coffey, Diane, and Dean Spears. 2017. *Where india goes: Abandoned toilets, stunted Development and the costs of caste.* Harper Collins.

Coffey, Diane, and Dean Spears. 2018. "Open defecation in rural India, 2015–2016: Levels and trends in NHFS-4." *Economic and Political Weekly*, 53(9): 10–13.

Cortes, Kalena E. 2004. "Are refugees different from economic immigrants? Some empirical evidence on the heterogeneity of immigrant groups in the United States." *The Review of Economics and Statistics*, 86(2): 465–480.

Cutler, David M., Adriana Lleras-Muney, and Tom Vogl. 2011. "Socioeconomic Status and Health: Dimensions and Mechanisms." In *Oxford Handbook of Health Economics.* , ed. Sherry Glied and Peter C. Smith. Oxford:Oxford University Press.

Deaton, Angus. 2013. *The great escape: health, wealth, and the origins of inequality.* Princeton University Press.

Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken. 2012. "Temperature Shocks and Economic Growth: Evidence from the Last Half Century." *American Economic Journal: Macroeconomics*, 4(3): 66–95.

Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken. 2014. "What Do We Learn from the Weather? The New Climate-Economy Literature." *Journal of Economic Literature*, 52(3): 740–798.

Deryugina, Tatyana, and Solomon M. Hsiang. 2014. "Does the Environment Still Matter? Daily Temperature and Income in the United States." NBER Working Paper.

Deschênes, Olivier, and Michael Greenstone. 2009. "Extreme Weather Events, Mortality, and Migration." *The Review of Economics and Statistics*, 91(4): 659–681.

Dillender, Marcus. 2017. "Climate Change and Occupational Health: Are There Limits to Our Ability to Adapt?" Working Paper.

Duflo, Esther, Rachel Glennerster, and Michael Kremer. 2008. "Using randomization in development economics research: a toolkit." In *Handbook of Development Economics.* , ed. T. Paul Schultz and John A. Strauss. Amsterdam and New York:Elsevier.

Dunne, John P., Ronald J. Stouffer, and Jasmin G. John. 2013. "Reductions in labour capacity from heat stress under climate warming." *Nature Climate Change*, 3: 563–566.

Edin, Per-Anders, Peter Fredriksson, and Olof Aslund. 2003. "Ethnic enclaves and the economic success of immigrants: evidence from a natural experiment." *Quarterly Journal of Economics*, 118(1): 329–357.

Floyd, Ife, and Liz Schott. 2015. "TANF cash benefits have fallen by more than 20 percent in most states and continue to erode." Center on Budget and Policy Priorities, Washington, D.C.

Frisbie, W Parker, Robert A Hummer, Daniel A Powers, Seung-Eun Song, and Starling G Pullum. 2010. "Race/ethnicity/nativity differentials and changes in cause-specific infant deaths in the context of declining infant mortality in the US: 1989–2001." *Population Research and Policy Review*, 29(3): 395–422.

GAO. 2011. "Refugee assistance: little is known about the effectiveness of different approaches for improving refugees' employment outcomes." Government Accountability Office, Washington, D.C.

GAO. 2012. "Refugee resettlement: greater consultation with community stakeholders could strengthen program." Government Accountability Office, Washington, D.C.

Geruso, Michael, and Dean Spears. 2018*a*. "Heat, Humidity, and Infant Mortality in the Developing World." NBER Working Paper 24870.

Geruso, Michael, and Dean Spears. 2018*b*. "Neighborhood sanitation and infant mortality." *American Economic Journal: Applied Economics*, 10(2): 125–62.

Ghosh, Arabinda, Aashish Gupta, and Dean Spears. 2014. "Are Children in West Bengal Shorter Than Children in Bangladesh?" *Economic & Political Weekly*, 48.

Graff Zivin, Joshua, and Matthew Neidell. 2012. "The impact of pollution on worker productivity." *American Economic Review*, 102(7): 3652–73.

Graff Zivin, Joshua, and Matthew Neidell. 2014. "Temperature and the Allocation of Time: Implications for Climate Change." *Journal of Labor Economics*, 32(1): 1–26.

Graff Zivin, Joshua S., Yingquan Song, Qu Tang, and Peng Zhang. 2018. "Temperature and High-Stakes Cognitive Performance: Evidence from the National College Entrance Examination in China." NBER Working Paper.

Greenstone, Michael, and Adam Looney. 2011. "Building America's job skills with effective workforce programs: a training strategy to raise wages and increase work opportunities." The Hamilton Project, the Brookings Institution, Washington, D.C.

Heal, Geoffrey, and Jisung Park. 2014. "Feeling the Heat: Temperature, Physiology, & the Wealth of Nations." NBER Working Paper.

Heal, Geoffrey, and Jisung Park. 2016. "Reflections-Temperature Stress and the Direct Impact of Climate Change: A Review of an Emerging Literature." *Review of Environmental Economics and Policy*, 10(2): 347–62.

Heckman, James J. 1976. "The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models." *Annals of Economic and Social Measurement*, 5: 475–492.

Heckman, James J., Seong Hyeok Moon, Rodrigo Pinto, Peter A. Savalyev, and Adam Yavitz. 2010. "The rate of return to the High/Scope Perry Preschool Program." *Journal of Public Economics*, 94(1-2): 114–128.

HHS. 2016. "Characteristics and financial circumstances of TANF recipients, fiscal year 2015." Office of Family Assistance, U.S. Department of Health and Human Services, Washington, D.C.

Ho, JY, and IT Elo. 2013. "The contribution of smoking to black-white differences in US mortality." *Demography*, 50(2): 545–68.

Holmstrom, Bengt, and Paul Milgrom. 1991*a*. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, & Organization*, 7: 24–52.

Holmstrom, Bengt, and Paul Milgrom. 1991*b*. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, & Organization*, 7, Special Issue: 24–53.

Hsiang, Solomon M. 2010. "Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America." *Proceedings of the National Academy of Sciences USA*, 107(5): 15367–72.

Hsiang, Solomon M., Marshall Burke, and Edward Miguel. 2013. "Quantifying the influence of climate on human conflict." *Science*, 341(6151).

Humphrey, Jean H. 2009. "Child undernutrition, tropical enteropathy, toilets, and handwashing." *The Lancet*, 374(9694): 1032–1035.

Johnston, David W, and Wang-Sheng Lee. 2011. "Explaining the female black-white obesity gap: a decomposition analysis of proximal causes." *Demography*, 48(4): 1429–1450.

Kaestner, Robert, and Neeraj Kaushal. 2005. "Immigrant and native responses to welfare reform." *Journal of Population Economics*, 18(1): 69–92.

Kjellstrom, Tord, and Jennifer Crowe. 2013. "Climate change, workplace heat exposure, and occupational health and productivity in Central America." *International Journal of Occupational and Environmental Health*, 27(3): 270–81.

Kjellstrom, Tord, David Briggs, Chris Freyberg, Bruno Lemke, Matthias Otto, and Olivia Hyatt. 2016*a*. "Heat, Human Performance, and Occupational Health: A Key Issue for the Assessment of Climate Change Impacts." *Annual Review of Public Health*, 37: 97–112.

Kjellstrom, Tord, Ingvar Holmer, and Bruno Lemke. 2009. "Workplace heat stress, health and productivity - an increasing challenge for low and middle-income countries during climate change." *Global Health Action*, 2.

Kjellstrom, Tord, Matthias Otto, Bruno Lemke, Olivia Hyatt, Dave Briggs, Chris Freyberg, and Lauren Lines. 2016*b*. "Climate Change and Labour: Impacts of Heat in the Workplace." United Nations Development Programme.

Lawson, Nicholas, and Dean Spears. 2016. "What doesn't kill you makes you poorer: Adult wages and early-life mortality in India." *Economics & Human Biology*, 21: 1–16.

Lemoine, Derek. 2017. "Expect Above Average Temperatures: Identifying the Economic Impacts of Climate Change." University of Arizona Working Paper 17-11.

LoPalo, Melissa. 2019. "The Effects of Cash Assistance on Refugee Outcomes." *Journal of Public Economics*, 170: 27–52.

LoPalo, Melissa, Diane Coffey, and Dean Spears. 2019. "The Consequences of Social Inequality for the Health and Development of India's Children: The Case of Caste, Sanitation, and Child Height." *Social Justice Research*, 1–16.

Marmot, Michael G, Geoffrey Rose, Martin Shipley, and Peter J Hamilton. 1978. "Employment grade and coronary heart disease in British civil servants." *Journal of Epidemiology & Community Health*, 32(4): 244–249.

Marmot, Michael G, Stephen Stansfeld, Chandra Patel, Fiona North, Jenny Head, Ian White, Eric Brunner, Amanda Feeney, and G Davey Smith. 1991. "Health inequalities among British civil servants: the Whitehall II study." *The Lancet*, 337(8754): 1387–1393.

Merel, Pierre, and Matthew Gammans. 2017. "Climate econometrics: Can the panel approach account for long-run adaptation?" Working paper.

Moffitt, Robert A. 2003. "The Temporary Assistance for Needy Families program." 291–363. University of Chicago Press.

MPI. 2015. "The Vietnamese diaspora in the United States." Migration Policy Institute, Washington, D.C.

ORR. 2012. "U.S. registered nursing re-licensing process." Office of Refugee Resettlement, Washington, D.C.

ORR. 2016. "Annual report to congress: FY2015." Office of Refugee Resettlement, Washington, D.C.

Park, Jisung. 2018. "Hot Temperature and High Stakes Exams: Evidence from New York City Public Schools." Working Paper.

Price, David J., and Jae Song. 2018. "The long-term effects of cash assistance." *Working Paper.*

Refugee Council. 2015. "The U.S. refugee admissions program: A presentation by refugee council USA for members of SCORR and ARHC." Refugee Council USA, Washington, D.C.

Schott, Liz, and Ladonna Pavetti. 2011. "Many states cutting TANF benefits harshly despite high unemployment and unprecedented need." Center on Budget and Policy Priorities, Washington, D.C.

Seppanen, Olli, William J Fisk, and QH Lei. 2006. "Effect of Temperature on Task Performance in an Office Environment." *Lawrence Berkeley National Laboratory.*

Sheffield, Justin, Gopi Goteti, and Eric F. Wood. 2006. "Development of a 50-Year High-Resolution Global Dataset of Meteorological Forcings for Land Surface Modeling." *Journal of Climate*, 19: 3088–3111.

Sherwood, Steven C., and Matthew Huber. 2010. "An adaptability limit to climate change due to heat stress." *Proceedings of the National Academy of Sciences*, 107(21): 9552–9555.

Spears, Dean. 2018. "Exposure to open defecation can account for the Indian enigma of child height." *Journal of Development Economics.*

Spears, Dean, and Amit Thorat. 2017. "The puzzle of open defecation in rural India: Evidence from a novel measure of caste attitudes in a nationally-representative survey." *Forthcoming in Economic Development and Cultural Change.*

Stevens, Andrew. 2016. "Temperature, Wages, and Agricultural Productivity." Working Paper.

Thorat, Amit, and Omkar Joshi. 2015. "The continuing practice of untouchability in India: Patterns and mitigating influences." India Human Development Study Working Paper 3.

UN. 2010. "State of the world's indigenous peoples. Press release."

Wilde, Elizabeth T., Zohn Rosen, Kenneth Coutch, and Peter A. Muennig. 2014. "Impact of welfare reform on mortality: An evaluation of the Connecticut Jobs First Program, a randomized controlled trial." *American Journal of Public Health*, 104(3): 534–538.

Zhang, Peng, Olivier Deschenes, Kyle C. Meng, and Junjie Zhang. 2018. "Temperature Effects on Productivity and Factor Reallocation: Evidence from a Half Million Chinese Manufacturing Plants." *Journal of Environmental Economics and Management*, 88: 1–17.